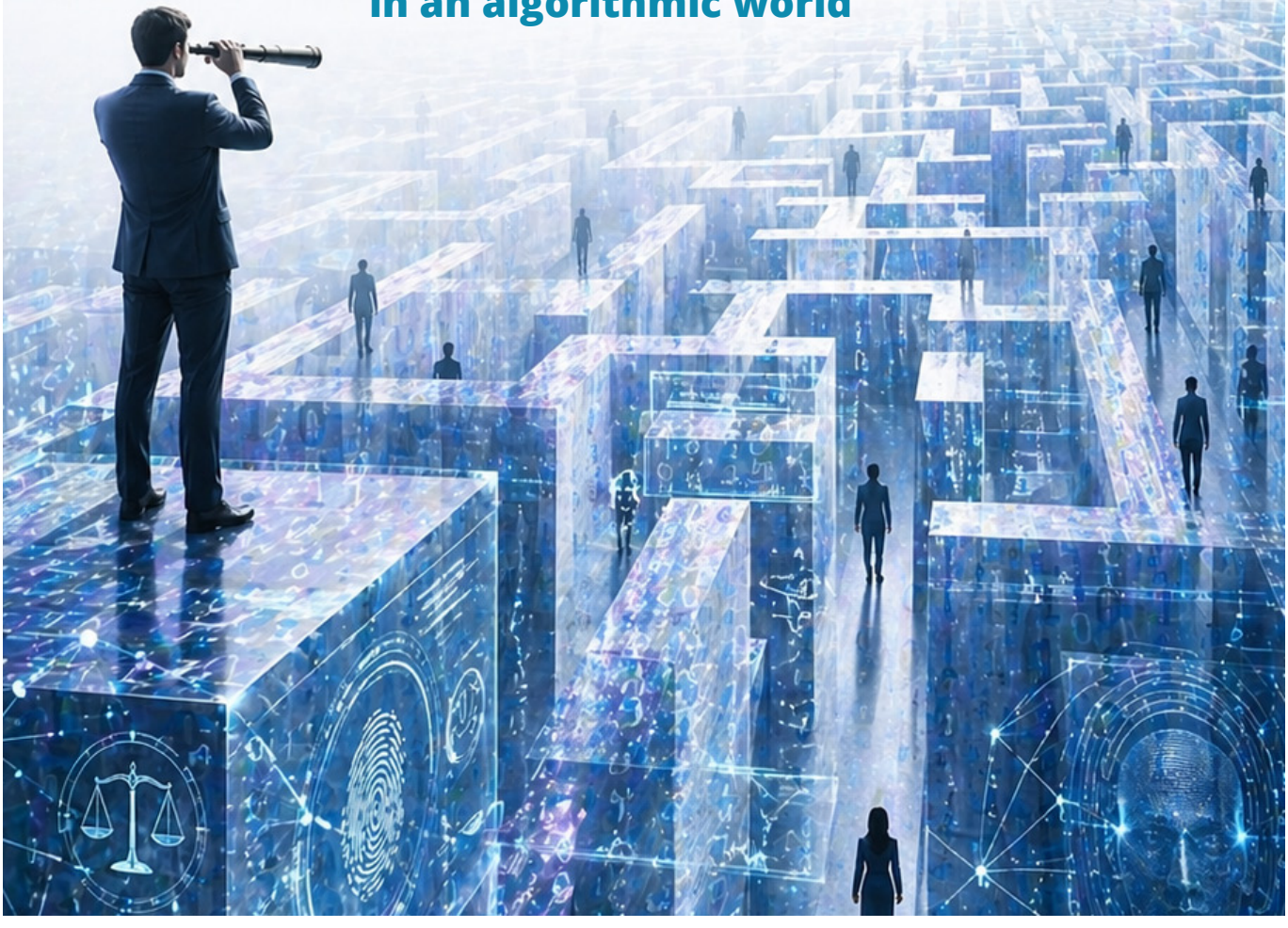
FREEDOM FROM FEAR
MAGAZINE

PROGRESS WITHOUT COMPASS

Rethinking crime, justice and rights in an algorithmic world



unicri
United Nations
Interregional Crime and Justice
Research Institute

GHENT
UNIVERSITY

Editorial Board

n°21

UNICRI

Ottavia Galuzzi
Marina Mazzini
Odhran McCarthy
Marco Musumeci
Leif Villadsen

Ghent University

Tom Vander Beken
Jelle Janssens
Noel Klima

Editor-in-Chief

Marina Mazzini

Editorial Team

Stephanie Briggs
Bryony Dennison
Ottavia Galuzzi
Marina Mazzini
Chris Plagnol

Graphic

Antonella Bologna

Cover image

AI-assisted visual development:
ChatGPT (OpenAI)

Disclaimer

The views expressed are those of the authors and do not necessarily reflect the views and positions of the United Nations. Authors are not responsible for the use that might be made of the information contained in this publication.

Contents of the publication may be quoted or reproduced, provided that the source of information is acknowledged.

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations and UNICRI, concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The mention of specific institutions, companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the Secretariat of the United Nations or UNICRI in preference to others of a similar nature that are not mentioned.

© UNICRI, MAY 2026

FREEDOM FROM FEAR
MAGAZINE

PROGRESS WITHOUT COMPASS

Rethinking crime, justice and rights in an algorithmic world

unicri
United Nations Interregional Crime and Justice Research Institute

GHENT UNIVERSITY

Contents

Between Daedalus and Icarus: Governing Risk in the Digital Labyrinth

by Leif Villadsen,
Acting Director of UNICRI

Today's global digital transformation is not merely technological; it is profoundly human, reshaping how we relate, decide, and understand the world. The myth of Daedalus and Icarus offers a compelling lens through which to interpret this moment. Daedalus, the father of Icarus, is both the architect of the labyrinth built to contain the Minotaur and the inventor of the wings designed to escape it. He embodies human ingenuity – the capacity to create systems of extraordinary sophistication to manage danger, but also the risk of becoming trapped within the very complexity we design. His role as a father adds a further layer of responsibility: he can provide tools and guidance, but cannot control how they are used. Icarus, by contrast, despite Daedalus' warning not to fly too close to the sun, represents the impulse to transcend limits, driven by ambition, possibility, and overconfidence. The closer humanity moves toward the metaphorical sun of limitless technological possibility, the more urgent become the questions of governance, ethics, and accountability. Together, Daedalus and Icarus capture a defining tension of our time – between creation and control, innovation and responsibility, freedom and risk.

This tension is not confined to myth. It is evident in modern history with the creation of the atomic bomb during the Manhattan Project – a turning point in human capability that demonstrated how scientific progress can outpace the frameworks needed to govern its consequences. Scientists, like Daedalus, harnessed extraordinary ingenuity to confront an existential threat, yet the power they unleashed raised enduring questions about responsibility, control, and the limits of human foresight.

Digital technologies are advancing at unprecedented speed, unlocking new opportunities for inclusion and development while simultaneously expanding the operational space for criminal actors. This dynamic reflects a fundamental asymmetry: technological capabilities grow exponentially, while societal understanding and institutional adaptation lag behind. As a result, threats – from opaque cybercriminal economies, growing overlaps among online harms and AI-enabled threats – are becoming more complex, less visible, and increasingly transnational, with impacts that extend beyond economic loss to affect trust, security, and fundamental rights.

At the same time, vulnerabilities are intensifying. Legal and regulatory frameworks remain largely confined within national borders, while cyber threats move seamlessly across them. Persistent challenges in attribution, accountability, and enforcement continue to expose structural gaps in the international legal architecture. Global efforts, including the new United Nations Convention against Cybercrime, mark important progress, but their effectiveness will depend on sustained political commitment and coordinated implementation.

This transformation is unfolding against a backdrop of escalating global tensions, persistent conflicts, and a visible erosion of multilateralism. As cooperation fragments, the capacity to address inherently transnational cyber threats is further constrained.

Addressing these challenges requires a balanced and comprehensive response. Strengthening normative frameworks is essential, but not sufficient. International cooperation must be deepened, and public–private partnerships reinforced, recognizing that much of the digital ecosystem lies beyond the direct reach of governments.

Young people stand at the centre of this transformation. They are among the most exposed to online harms, yet also among the most capable of shaping responses. As digital natives, they operate in environments defined by anonymity, algorithmic influence, and blurred boundaries between physical and virtual realities. Strengthening digital awareness, critical thinking, and behavioural understanding is therefore not only protective – it is strategic. Investing in digital literacy and positioning youth as active contributors to cybersecurity and crime prevention offers a pathway towards more sustainable solutions.

This issue of F3 brings together diverse perspectives that reflect the multifaceted nature of cyberspace today. From legal analysis to behavioural insights, from technological innovation to community-based initiatives, the contributions highlight both the complexity of the challenges and the breadth of responses required.

No single actor can address these challenges alone. A collective effort – grounded in shared responsibility, informed by knowledge, and guided by a commitment to justice and human rights – remains essential. Only through such an approach can technological progress serve not as a vector of harm, but as a foundation for safer and more equitable societies.

Yet, without clarity of direction, even collective efforts risk falling short. As depicted in The Parable of the Blind by Pieter Bruegel il Vecchio, a line of individuals advances across uncertain ground, each relying on the next, unaware of the risks ahead. It is a stark reminder that progress without vision does not merely fail – it amplifies vulnerability. Ensuring that awareness, responsibility, and cooperation guide our steps is therefore not optional, but essential to building safer, more just and resilient societies.

Online harms affect youth: growing threats and integrated responses

by **Human Digital**

> **"Gore sites[1] and online communities sharing graphic violent content are largely overlooked in online safety regulation and debate, despite being spaces where violent extremism, violent misogyny and a non-ideological interest in graphic violence converge."**

To date, easy access, limited age verification and an absence of regulation have left children – and young people – at significant risk of harm from these spaces. Addressing this risk requires an integrated response across child safety regulation, public health prevention and national security spheres. Responses could vary from policy and regulatory sanctions on those making 'gore' available, to platform engagement, content referral, and targeted disruption of illegal activity by operational or strategic communication means.

How real is the threat?

Although they are not usually the reason graphic violent content is produced in the first place, Gore sites are crucial to the persistence and propagation of harmful content online – and they enable more people to find, watch and share it. A growing body of evidence has highlighted the risk that unrestricted access to graphic violent videos and imagery poses to young people.[2]

Based on research carried out by Human Digital, Gore sites host significant volumes of unmoderated terrorist propaganda, bloody violence, abuse imagery and extreme pornography. Many of these sites are intentionally designed for the download

[1] "Gore sites" are online platforms that distribute graphic content, including images and videos, depicting real-life violence, injury, or death, often with limited or no moderation. In some cases, these platforms may also host and disseminate authentic audiovisual footage produced by violent extremist actors or associated with other forms of serious criminal activity, thereby amplifying the reach and impact of such content.

[2] https://www.counterterrorism.police.uk/ctp-joins-five-eyes-partners-in-calling-for-whole-society-action-to-fight-growing-threat-to-children-posed-by-online-extremism/

“

Strikingly, at the time this research was conducted, little to no content was prohibited by Gore sites. No sites had any effective age verification, paywalls or content warnings restricting access to children

or onward sharing of such material on social media. Despite online legislation in the EU, UK and other regions requiring platforms to implement baseline safety mechanisms, these sites fail to provide the types of barriers employed by tech platforms to protect their users – especially children.

Archives of cross-harm content can be easily found and navigated by curated lists, including tags such as "ISIS" and categories such as "terrorism" or "war", that guide users directly to terrorist content.

Whilst themes vary across Gore sites, almost all feature what Human Digital calls "B.A.D. Content": bloody violence, abuse and death. Graphic footage of fatal accidents, suicides, stabbings and shootings, self-mutilation, terrorist propaganda, animal abuse, and drug cartel murder videos are curated and shared specifically for their graphic nature. Advertisements for pornography sites are frequently observed via banner ads and pop-ups, and Human Digital have identified web traffic trends of people moving directly from pornography sites to Gore sites.

To analyse the accessibility of B.A.D Content, Human Digital assessed 24 of the most globally visited gore-related websites. The subsequent "Gore and Violent Extremism" report,[3] published by VOX-Pol, found that search engines return millions of results for gore-related keywords, revealing thousands of terrorist-made videos, and evidence of extreme pornography, including mutilation and bestiality.

"Analysis of web traffic to these 24 websites shows an average of over 24 million total visits per month globally in 2026."

Strikingly, at the time this research was conducted, little to no content was prohibited by Gore sites. No sites had any effective age verification, pay-walls or content warnings restricting access to children. These findings provide evidence that Gore sites are currently overlooked by many law enforcement organisations, online regulators and industry moderation policies. Today, 13 of the sites still have no barriers to access, 2 required a simple 'Yes/No' over-18 verification question, 1 asked the visitor to decide whether they wanted to click through to the content, 6 were no longer active, 1 was inaccessible in the UK and 1 required a sign-up account for access. This data speaks to both the continued harm, volatility and resilience of the ecosystem.

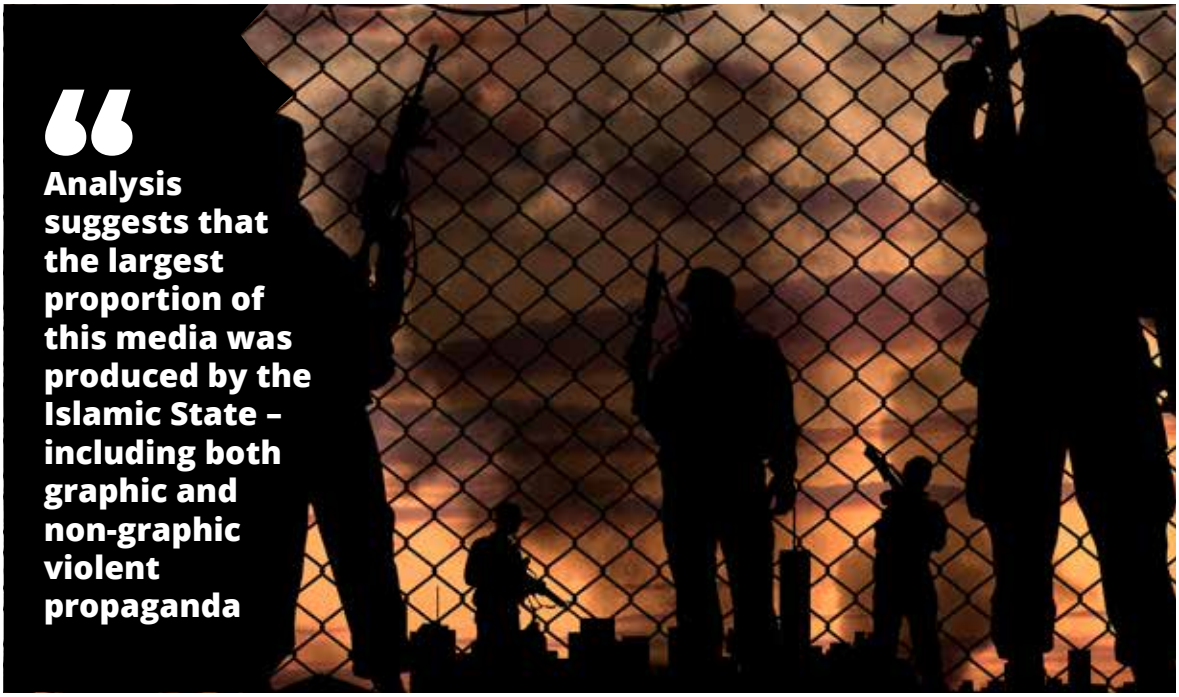
"This ecosystem of these sites has the potential to cause real-world harm, ranging from national security threats to public safety hazards."

As such, public health responses to support victims and young people exposed to Gore content are increasingly important.

What national security risks do Gore sites present?

In 2024, searches for terrorist material on 24 Gore sites found over 12,000 non-unique examples of relevant content. Analysis suggests that the largest proportion of this media was produced by the Islamic State – including both graphic and non-graphic violent propaganda. Hundreds of examples of the Christchurch, Buffalo and Halle attack livestreams were also found, accruing hundreds of thousands of views in total.

[3] https://voxpol.eu/wp-content/uploads/2025/07/DCUPN0751-Gore-Extremism-WEB-250704.pdf



“Analysis suggests that the largest proportion of this media was produced by the Islamic State – including both graphic and non-graphic violent propaganda

These videos are often uploaded by gore enthusiasts, creating repositories of historic and recent terrorist material. While Human Digital researchers uncovered only limited evidence that designated terrorist organisations were actively exploiting Gore sites, the presence of so many violent videos create pathways through which vulnerable users could be exposed to radicalising content.

In addition to prominent terrorist groups, the researchers have observed across social media how violent extremist groups and communities are sharing an increasingly diverse range of harmful content online, with members – often young – whose worldviews do not fit into established ideological categories. In the UK, since 2018, the Prevent Strategy has reported a rise in referrals of young people who do not adhere to a clear or coherent ideology. Between 2023 and 2024, 66% of Prevent referrals belonged to categories which would have previously been subsumed under Mixed, Unclear and Unstable; 32% of those were under the age of 15.[4] In some of these cases, a fascination with violence displaces a fixed ideology.

Incidents such as the 2024 Southport attack demonstrate the potential risk posed by individuals who display no clear ideological affiliation but nonetheless maintain a fascination with mass violence.[5] Sir Adrian Fulford of the Southport Inquiry stated, “that the degrading, violent and misogynistic material which Axel Rudakubana was viewing online contributed to – and fed – his already unhealthy fascination with violence.”[6]

[4] The category “Mixed, Unstable and Unclear” was broken down into 7 component categories in March 2024; https://www.gov.uk/government/collections/individuals-referred-to-and-supported-through-the-prevent-programme-statistics.

[5] Jonathan Hall KC, Independent Reviewer of Terrorism Legislation, identified 10 cases involving extreme violence that lacked a clear ideological driver in Independent Review on Classification of Extreme Violence Used in Southport Attack on 29 July 2024 (13 March 2025)

[6] See https://www.southport.public-inquiry.uk/

> **"Several recent UK terrorism convictions demonstrate an interest in Gore, often from a very early age and blended with extreme right-wing ideology, making it difficult to parse out the influence of ideology from violence fascination.[7]"**

Following the shooting of Charlie Kirk, one popular Gore site crashed due to web traffic, which the site explained to be a result of people searching for the assassination video and creating new accounts. This increases exposure to ideological content for those visitors, regardless of their intent in seeking out gore.

How is this a public safety issue?

There have been multiple recent cases globally of individuals carrying out violent acts after engaging with graphic violent content on Gore sites, including what the National Centre of Missing and Exploited Children describe as "cross-genre" content.[8] Similarly, since 2022 Human Digital has tracked the increasing prevalence of graphic violent and violent misogynistic content within and adjacent to violent, extremist online ecosystems. Often, this content exists as embedded videos displaying Gore site watermarks, including content related to the Com network – "online networks of predominantly teenage boys, dedicated to inflicting harm and committing a range of criminality"[9] - and an orientation towards a fascination in non-ideological violence and ultra-sadism. Gore sites host content referenced within these harmful groups and provide content stores through which abusive content can be shared and normalised. Accounts on select gore sites have also been implicated in direct off-linking to Com-affiliated group chats on messenger platforms, hence Gore sites acting as both a content repository and gateway into increasingly harmful and abusive online ecosystems.

Individuals emulating – or encouraging others to engage in – such violence have a clear and acute impact on their victims. These victims are often members of vulnerable communities, such as online discussion groups dedicated to eating disorders, depression, or anxiety, and are often young girls. In some cases, the only escape for a victim is to become a perpetrator themselves, sustaining a cycle of extreme violent content production and consumption.

In another example, a popular incel forum held over 2,000 posts directing users to gore-related websites, some of which advertised purchasable folders of videos of women being murdered. There is an acute risk to women and girls within many of the communities sharing links to Gore sites, both in terms of the victims within the content and the targets of violent attacks.

What might an integrated response look like?

The threat posed by the converging harm sets visible on Gore sites – and across social media – presents short - and long-term public safety and national security risks that demand an internationally aligned response from UN Member States, law enforcement agencies, and regulators. An effective shared strategic vision would set preventative and responsive objectives into a coherent

[7] See Felix Winter, 2025 at https://www.bbc.co.uk/news/articles/clyj2g5l1g2o and Vincent Charlon, 2023 at https://www.bbc.co.uk/news/articles/ceke1el177xo

[8] https://newsroom.co.nz/2024/03/27/nz-teen-arrested-with-trove-of-violent-sexual-and-terrorist-content/

[9] https://www.nationalcrimeagency.gov.uk/news/sadistic-online-harm-groups-putting-people-at-unprecedented-risk-warns-the-nca

“There have been multiple recent cases globally of individuals carrying out violent acts after engaging with graphic violent content on Gore sites

framework combining policy and regulation, platform engagement, content referral, and targeted disruption of illegal activity. Such a framework should avoid fragmented cross-sector action and the duplication of effort, seen within the “whack-a-mole” responses to violent extremism on mainstream social media. Recent studies have shown that an outcome of pornography sites requiring more robust age verification when accessed by UK IP addresses has been that many in the UK now watch content on ‘unregulated’ pornography sites. This relates to convergent harm spaces such as Gore sites which advertise and share web traffic referrals with pornography sites, including those that are unregulated and unmoderated and contain violent, harmful sexual material.[10] In addition to direct engagement with and regulation of Gore site owners, both search engines and mainstream social media platforms could be tasked with limiting the discoverability of these Gore sites, particularly by youth.

“Technology and data analytics will be central to enabling an internationally coordinated response to Gore sites.”

Stakeholders require insights to understand both how and why young people encounter and engage with such content, to help guide the formulation of harmonised policies, strengthen cross-border regulatory cooperation, and ensure that interventions are both targeted and proportionate.

[10] https://www.independent.co.uk/news/uk/home-news/unregulated-sites-porn-uk-age-verification-b2895231.html

This intelligence can tailor communications campaigns to deter or direct young people away from harmful content, including educating them on the legal, ethical and psychosocial risks it can pose. As an online harm set,

> **"research has highlighted the potential long-term impact on individuals who view extreme violence,[11] including chronic and long-term mental health implications."**

Unlike the small number of violence-inspired attackers whose actions gain widespread coverage, the chronic and cumulative impact of Gore sites on a larger group of vulnerable viewers can go unnoticed in wider policy discussion. There is also increasing evidence of the impact of vicarious trauma and other harms among professionals working on Gore content, caused by viewing graphic and violent material.[12]

A comprehensive approach to the impact of Gore sites and graphic violent content online should include clear guidance on how to work with victims, perpetrators, and viewers of violent content. Beyond digital responses, educational programmes could be delivered for intervention practitioners so they can better support children and young people and build an evidence base of successful approaches to support and protect children who become victims of this new, but growing, cross-harm threat. Significantly, the long-term effect of exposure to B.A.D content will need to be monitored. It is currently not known what impact the exposure to violent content online from a young age will have on the citizens of UN member states, including its future police officers, military personnel, teachers, and politicians.

Ultimately, innovative and informed responses are urgently needed to help those children and young people already exposed to Gore sites and protect them from exposure to this content in the future.

About Human Digital

This article was produced by researchers at **Human Digital,** a division within M+C Saatchi World Services who develop data-led understanding and recommendations to combat online harms, including violent extremism, violence fixation, misinformation, and the exploitation of emergent technologies, for example, AI, cryptocurrency and Web3. The team consists of analysts, data scientists and engineers working collaboratively with subject specialists from across academic disciplines.

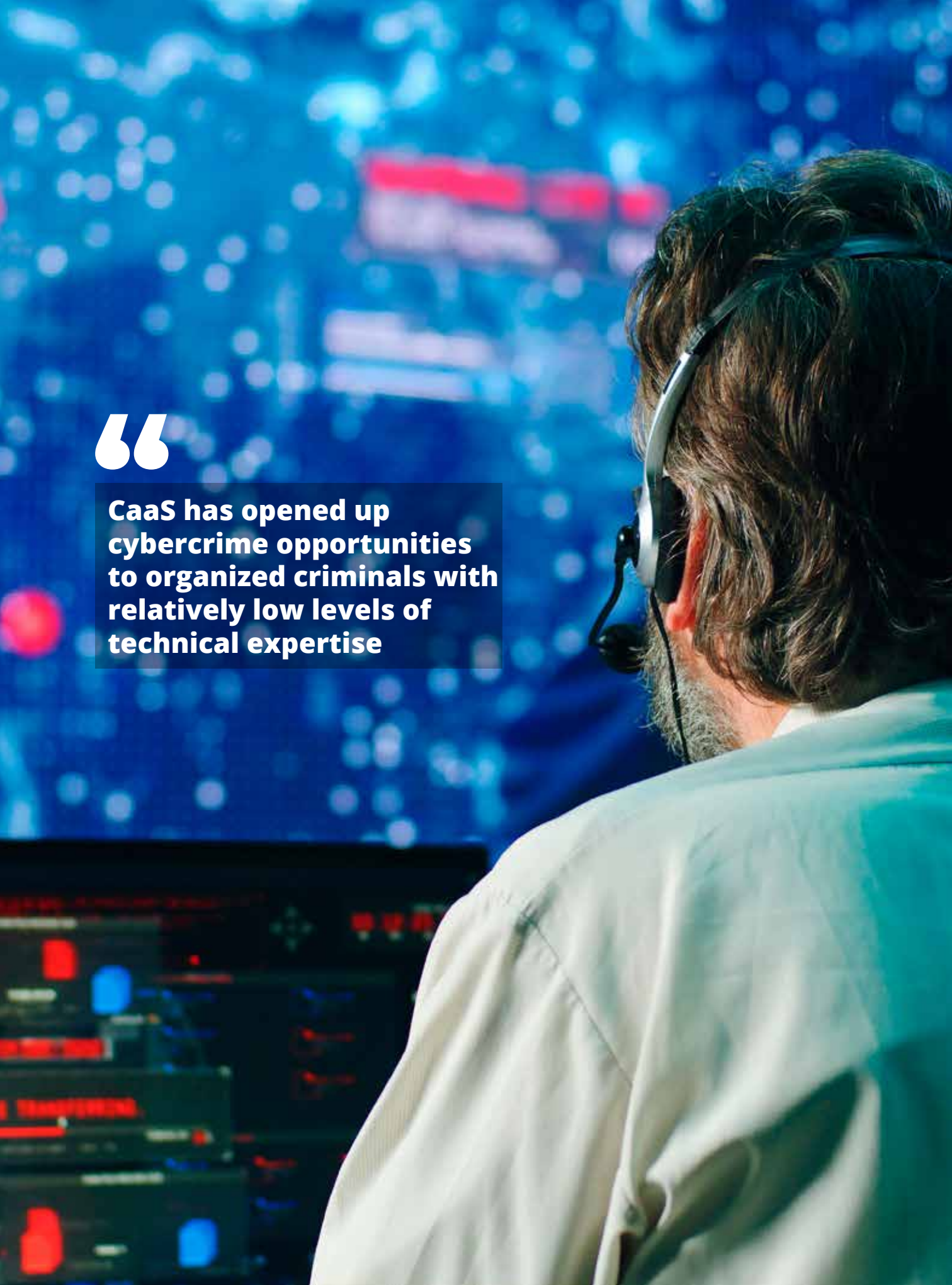
[11] Nicklin LL, Swain E, Lloyd J. Reactions to Unsolicited Violent, and Sexual, Explicit Media Content Shared over Social Media: Gender Differences and Links with Prior Exposure. Int J Environ Res Public Health. 2020 Jun 16;17(12):4296. doi: 10.3390/ijerph17124296. PMID: 32560142; PMCID: PMC7345319.

[12] https://www.swansea.ac.uk/law/cytrec/projects/reassure/

United Nations
Interregional Crime and Justice
Research Institute

Sport can no longer be considered a luxury within any society but is rather an important investment in the present and future, particularly in developing countries.

Summer schools 2026



> **CaaS has opened up cybercrime opportunities to organized criminals with relatively low levels of technical expertise**

Tracking the unseen: measuring the financial impact of cybercrime

by **Seán Doyle** and **Giulia Moschetta**

As the global economy digitizes, cybercriminals have ignored any and all attempts at deglobalization.[1] Their technical and personal networks continue to expand across regions, making it increasingly difficult to disrupt online criminal activities.

In the United States of America alone, the Federal Bureau of Investigation (FBI) reports that victim complaints exceeded USD 16 billion in 2024.[2] Research from the Global Anti-Scam Alliance suggests that scammers stole more than USD 442 billion worldwide in 2024-25.[3]

Changing structure and culture of cybercriminal groups

Cybercriminal activities are shifting – metaphorically and literally – into the sphere of traditional transnational organized crime groups.[4] They build on a trend spanning more than a decade in which technically sophisticated cybercriminals have withdrawn from roles as the direct executors of primary offences, such as a ransomware attack, to become developers and providers of criminal services. Their activities range from malware creation to ransom negotiation, and are known collectively as "Cybercrime-as-a-Service" (CaaS).

CaaS has opened up cybercrime opportunities to organized criminals with relatively low levels of technical expertise. Apart from increasing the pool of cybercriminals and the volume of attacks, this marked a turning point in the culture of cybercriminals: from a group that generally tried to bypass law enforcement attention by avoiding crimes that directly resulted in death or physical harm, to a culture where active disruption of key services, including disruption leading to death[5], such as by shutting down a hospital, has become common.

1 World Economic Forum, Global Cybersecurity Outlook 2026, January 2026.

2 Federal Bureau of Investigation Internet Crime Report 2024.

3 Global Anti-Scam Alliance, Global State of Scams 2025 Report.

4 As one example, see the movement of the Black Axe criminal network and cult into cyber-enabled fraud, "United States Attorney's Office, District of New Jersey, Prominent Leader of Black Axe Extradited to United States for Conspiring to Engage in Internet Scams and Money Laundering, 16 December 2024.

5 BBC, Ransomware attack contributed to patient's death", 25 June 2025.

New technologies, new markets, new victims

Technologies such as generative artificial intelligence (GenAI) and AI-supported deepfakes are making it easier for cybercriminals to enter new markets at low cost, by supporting translation of social engineering tricks into impersonations that are culturally credible in multiple countries and support initial access into the victim's systems (if an organization) or trust (if an individual).[6]

These technologies allow criminal groups who might previously have focused on populations with widely spoken languages like English, Mandarin, French, Russian, Arabic and Spanish, to target new populations of victims, in countries where traditional cybercriminal social engineering tactics had previously not been successful.

What is the impact of cybercrime on societal security?

Cybercrime's growth requires urgent attention. The funding it provides to organized crime creates an ongoing risk of state capture in some jurisdictions,[7] and challenges in countering it can exacerbate tensions between states where victims are concentrated and states from which cybercriminals operate.[8]

The culture of cybercrime is also changing – for the worse. This is occurring globally, with some of

the starkest examples in Anglophone regions where there has been an increase in "cyber thuggery",[9] as a reference to the rising risk appetite of cybercriminal groups with a lower age profile, from mid-20s to late teens.

An example of this is "the Com" a shifting community of more than a thousand – mostly young - hackers whose activities, according to the US FBI, are spread across cyber-enabled and violent crime, like "swatting/hoax threats, extortion/sextortion of minors, production and distribution of child sexual abuse material, violent crime, and various types of cyber crimes."[10]

---

6 Europol, Steal, Deal and Repeat - How Cybercriminals Trade and Exploit Your Data – *Internet Organised Crime Threat Assessment, Publications Office of the European Union, Luxembourg, 2025.*

7 Assessment based on multiple sources including, Amnesty International, "I was someone else's property": Slavery, Human Trafficking and Torture in Cambodia's Scam Compounds, 26 June 2025.

8 International Crisis Group, Border Dispute with Cambodia Sparks Political Disarray in Thailand, 1 July 2025. See also tensions between Thailand and Myanmar, CNN Power Cut to Site of Global, Billion-Dollar Scam Industry. But Will It Halt the Swindling?, 5 February 2025.

9 Martin, Ciaran, Thugs, Thieves and Other Threats: How What We Need to Worry About in Cyberspace is Changing, 16 September 2025.

10 FBI Public Service Announcement I-072325-3-PSA, The Com: Theft, Extortion and Violence are a Rising Threat to Youth Online, 23 July 2025.

> **"The FBI describes the motivations behind the crimes as varied, but extending beyond pure financial gain to include "retaliation, ideology, sexual gratification, and notoriety."[11]"**

Similar to the script kiddies[12] more prevalent in the early 2000s and largely motivated by entertainment or notoriety, today's emerging cybercriminal groups often rely on relatively unsophisticated social engineering techniques to gain access to systems.[13] Yet, the level of harm they can inflict is now significantly greater, due to the deep interconnectivity and interdependence of digital supply chains. Groups like Scattered Spider appear to be connected with online violent subcultures and have been cited as the source of cyber-attacks on manufacturing giants such as Jaguar Land Rover and retail chain M&S in the United Kingdom.[14] The shift we are witnessing is not so much in the choice of targets as in the nature of the attacks themselves; contemporary attacks increasingly aim to disrupt or halt the operations of critical services such as manufacturing, healthcare, and food supply.

Is there an impact on state security?

Governments may already be adopting similar approaches, with the focus of cyber-attacks moving from espionage to coercion. For example, in August 2025, Norway blamed an adversarial state, or attackers in sympathy with it, for a cyber-attack that targeted the operations of a dam. The cyber-attack released large volumes of water over several hours and was interpreted by the Norwegians as an attempt to demonstrate destructive capability without acting on it.[15] Otherwise put, the aim of the attack was seen as supporting the attacker's political goals by intimidation.

We are getting better at disrupting cybercriminal networks

Law enforcement is steadily improving in its ability to coordinate cybercrime disruption across borders, enhancing the insights it gains into cybercriminal activities by working in partnership with experts in the private sector.

Cross-border operations, like the joint INTERPOL and Afripol Operation Serengeti 2.0 in 2025, supported by actionable insights into cybercrime from stakeholders outside law enforcement authorities, are now more frequent, more sophisticated and increasingly effective in dismantling international criminal networks.[16] Europol's European Cybercrime Centre (EC3) regularly relies on its network of operational partnerships with Internet security providers, financial services and telecommunication providers to increase the impact of its operations and investigative efforts across European and extra-European Member States. [17]

In Operation Serengeti 2.0, INTERPOL acted as a hub for sharing expertise and threat assessments between law enforcement, the private sector and non-governmental collaborations like the World Economic Forum-hosted Cybercrime Atlas, which

[11] *Ibid.*
[12] Okta, Script Kiddies and Skiddies: Identifying Unskilled Hackers, 2 September 2024.
[13] Sans, Defending Against Scattered Spider and the Com with Cybercrime Intelligence, 4 July 2024.
[14] BBC, Tidy, Joe "M&S Hacker Claims to be Behind Jaguar Land Rover Cyber Attack, 3 September 2025.
[15] Politico.eu, Russian Hackers Took Control of Norwegian Dam, Police Chief Says, 13 August 2025.
[16] Doyle, Seán and Umansky, Natalia, Cybercrime Is Borderless. This Global Bust Shows Law Enforcement Can Be Too, 27 August 2025.
[17] For example, Europol and Microsoft Disrupt World's Largest Infostealer, Lumma, 21 May 2025.

is a hub for the generation and distribution of private sector insights into cybercriminal networks.[18] The principles for developing sustained and effective operational partnerships to counter cybercrime[19] are now well-known, repeatable and integrated into international law enforcement trainings.

The accelerating scale and impact of cyber-enabled fraud call for systemic defence

Phishing and cyber-enabled fraud are a growing global threat to users, consumers, organizations, and countries. According to Global Cybersecurity Outlook 2026 survey data, 73% of respondents reported that they or someone in their network had been personally affected by cyber-enabled fraud in 2025.[20] Recognizing the need to rebalance responsibility for cybersecurity, the World Economic Forum's Partnership Against Cybercrime – together with the Institute for Security and Technology (IST) – have published the white paper "Fighting Cyber-Enabled Fraud: A Systemic Defence Approach."[21] Building on the Partnership's progress in fostering public–private operational collaboration, this white paper seeks to advance shared responsibility across the digital ecosystem by stimulating coordinated action and required policy reforms, led by key stakeholders.

This Systemic Defence framework calls on stakeholders to act across 3 pillars:

1. **Prevention** by embedding safeguards at the foundational layers of the Internet to reduce threat actors' ability to acquire, build, or operate digital infrastructure for malicious purposes.

Actions could include strengthening risk-based due diligence and oversight in domain registration and hosting, to detect and prevent misuse long before harm occurs.

2. **Protection** by ensuring safety by default, such as by embedding scalable solutions for consumer-facing services to shield users from phishing and cyber-enabled fraud. Governments can accelerate adoption and impact through national coordination hubs, enabling regulation, and targeted incentives.

[18] World Economic Forum, Cybercrime Atlas Annual Impact Report 2025, October 2025.
[19] World Economic Forum, Disrupting Cybercrime Networks: A Collaboration Framework, November 2024.
[20] World Economic Forum, Global Cybersecurity Outlook 2026, January 2026.
[21] World Economic Forum. (2025) Fighting Cyber-Enabled Fraud: A Systemic Defence Approach. November 2025.

3. **Mitigation** by improving ecosystem-wide capability to identify abuse, enable effective reporting, and share actionable signals, while also supporting rapid response, to both takedown malicious activities from upstream infrastructure and update downstream protection efforts.

With phishing and cyber-enabled fraud growing at an alarming rate – driven in part by cybercriminals' use of AI – stronger defences must be built with appropriate safeguards to ensure that legitimate users are protected from criminal abuse.

About the Authors

Seán Doyle leads the Cybercrime Atlas, an open-source investigations and applied research collaboration hosted at the World Economic Forum in Geneva, Switzerland. He has been at the World Economic Forum's Centre for Cybersecurity since its launch in 2018, leading the Forum's research into how organization's respond to cyberthreats while also building collaborations in cross-sector intelligence sharing. Before moving into cyber, Seán worked in cross-border financial crime intelligence and asset tracing. His career started as an Armed Forces Analyst focusing on Eastern Europe and the Former Soviet states.

Giulia Moschetta leads the Partnership against Cybercrime, an initiative fostering public–private collaboration to combat cybercrime, at the World Economic Forum's Centre for Cybersecurity. She also leads the Global Cybersecurity Outlook, the Forum's annual flagship report on key cybersecurity trends. She previously held roles at NATO and within the Italian government.

The cyberpsychology of AI-enabled cybercrime: human factors, emerging threats, and building resilience

by **Professor Mary Aiken**

Introduction

Cybercrime is not merely "crime with computers," in practice, it is a behavioural problem unfolding in technical environments. Offenders exploit platform design, social cues, and cognitive shortcuts at scale; victims face hyper-personalized attacks, with minimal costs to the attackers. From mass-scale cyber fraud to deepfake-enabled impersonation, offenders leverage platform design, human decision-making, cognitive biases, and social dynamics as much as they exploit code. The result is a threat landscape that grows in scale and harm, particularly as tools for committing cybercrime become increasingly easy to acquire and use. In 2024, the Federal Bureau of Investigation's (FBI) Internet Crime Complaint Center (IC3) reported 859,532 complaints and losses of $16.6 billion, a 33% jump from 2023, illustrating how effectively offenders are weaponizing persuasion online and underscoring the velocity and cost of cyberenabled fraud.[1]

Cyber fraud and scams continue to escalate on a global scale. According to the 2024 Global Anti-Scam Alliance (GASA) Report, international scammers collectively defrauded victims of more than USD 1 trillion over a twelve-month period.[2] The subsequent 2025 GASA Global Scam Report, which surveyed over 46,000 adults across 42 markets, revealed that 57% of respondents had encountered a scam within the past year, while 23% reported financial losses as a result.

A recent INTERPOL report (2025) has highlighted a sharp rise in cybercrime in Africa, stating that cybercrime now accounts for more than 30 per cent of all reported crime in Western and Eastern Africa. Online scams, ransomware, business email compromise and digital sextortion were the most reported cyber threats.[3] Of particular concern in a global context is the trafficking of victims to work in cyber fraud and scam centres; a recent crime trend report released by INTERPOL highlighted

1 FBI, "Internet Crime Report 2024", Internet Crime Complaint Center.
2 GASA, "Global State of Scams Report 2024".
3 INTERPOL, "New INTERPOL report warns of sharp rise in cybercrime in Africa", June 2025.

Offenders exploit platform design, social cues, and cognitive shortcuts at scale; victims face hyper-personalized attacks, with minimal costs to the attackers

that victims have been trafficked into criminality from more than 60 countries around the world.[4] This article outlines a cyberpsychology perspective on cybercriminal offending and victimization online. It connects to operational and policy debates, concluding with evidenceinformed recommendations for law enforcement, policymakers, industry, and civil society.

Global reports highlight the financial impact, with billions lost annually; however, there is also the human cost, ranging from the targeting of seniors to youth drawn into cybercrime through curiosity, status-seeking, and peer influence. A cyberpsychology perspective reframes this challenge:

> **"Prevention must target environments and decision points, not just awareness-raising initiatives."**

This article examines the definitions, drivers, and psychological dynamics of online offending and victimization, and presents six evidence-informed recommendations for law enforcement, policymakers, industry, and civil society. The aim is to embed human-centred resilience into systems and governance - creating a safer and more secure cyberspace for all.

Conceptualizing cybercrime: definitions, typologies and taxonomies

A persistent challenge for strategy, measurement, and resourcing is that there remains no single, universally accepted definition of cybercrime. A recent comprehensive review of academic and grey literature found significant variation in terminology, classifications, typologies, and taxonomies – variations that reverberate through law, policy, enforcement cooperation, and research comparability.[5] The authors concluded that "developing a clear conceptualization of cybercrime is needed not only to delineate the problem, but also for estimating the impact of cybercrime on society, and developing effective legal and policy responses".

[4] INTERPOL, "INTERPOL releases new information on globalization of scam centres", June 2025.

[5] Phillips, Kirsty, Julia C. Davidson, Ruby R. Farr, Christine Burkhardt, Stefano Caneppele, and Mary P. Aiken, "Conceptualizing Cybercrime: Definitions, Typologies and Taxonomies" *Forensic Sciences 2, no. 2: 379-398, 2022.*

To effectively classify current and emerging cybercriminal behaviours, a more comprehensive classification framework is required, notably one that is compatible with international and national legislation and policies. Internationally, the Budapest Convention on Cybercrime (Council of Europe, European Treaty Series No. 185) remains one of the principal binding frameworks for aligning offences and enabling cross-border cooperation and electronic evidence handling. Notably, it represents the only globally recognized agreement on cybercrime. More recently, the United Nations Convention against Cybercrime[6] has been adopted to establish a broader global framework for international cooperation, capacity building, and the harmonization of certain cybercrime-related offences, although its scope and implementation remain the subject of ongoing policy and legal debate. However, technological innovations, such as AI-assisted fraud and sophisticated privacy-enhancing technologies, will continue to drive definitional and procedural debates.

Human and technical drivers: why crime scales online

From a cyberpsychology standpoint, classic mechanisms such as the Online Disinhibition Effect, anonymity and the minimization of authority online[7] combine with the affordances of digital platforms to lower restraint and accelerate harm. Prevention, therefore, must target environments and decision points, not only individual and societal awareness.

The overall cybercrime economy has professionalized. Crime-as-a-Service (CaaS) markets offer intrusion kits, access brokering, money laundering, and even forms of "customer support," thereby reducing the skill threshold and increasing productivity for offenders. Europol's Internet Organized Crime Threat Assessment (IOCTA) report outlines how this division of labour, coupled with scalable targeting, reshapes opportunity structures online.[8] However, technical supply is only half the story. EU research (e.g., CC-DRIVER) emphasizes human factors, including curiosity, risk-taking, status-seeking, and peer influence, as pivotal drivers of juvenile cyber-delinquency and cybercriminality. A nine-country youth survey (N = 8,000 participants, aged 16 to 19) found that just under half (47.76%, N = 3,808) reported engaging in criminal behaviour online in the previous year. Prevalence rates for individual behaviours ranged from 1 in 8 engaging in money muling and laundering, to 1 in 11 engaging in cyber fraud.[9] The findings underscore how online risk-taking correlates with offending and how tailored, evidence-based interventions can redirect trajectories.

Entry points for young people into cybercrime are abundant, cyber behavioural drivers can be conceptualized as follows:

- **Exposure and normalization**: Dark-web marketplaces and forums put Crime-as-a-Service tools, stolen data, and "starter" services within easy reach, socializing and normalizing illicit experimentation for tech-savvy and curious youth.
- **Lowered restraint**: Perceived anonymity online and encryption reduce psychological inhibition and risk perception, enabling neutralizations ("no real victim") that can make offending feel acceptable.

6 UNODC, "United Nations Convention against Cybercrime; Strengthening International Cooperation for Combating Certain Crimes Committed by Means of Information and Communications Technology Systems and for the Sharing of Evidence in Electronic Form of Serious Crimes", 2024.

7 Mary P. Aiken, "Introduction to cyberpsychology", Global Cybersecurity Forum, September 2024.

8 Europol, "Internet Organised Crime Threat Assessment (IOCTA) 2024".

9 Julia Davidson, Mary P. Aiken, Kirsty Phillips and Ruby Farr, "CC-DRIVER 2022 Research Report".

- **Status and rewards**: Reputation systems, mentorship from seasoned offenders and instant crypto-monetization create powerful statusandreward reinforcement loops that can escalate experimentation into organised participation.

The UK's 'Cyber Choices' programme was created to help individuals make informed choices and to use their cyber skills legally, illustrating diversion in practice by clarifying legal boundaries, engaging at early touchpoints and redirecting juvenile talent into legitimate cybersecurity pathways through competitions and mentoring. Such approaches treat youth involvement as a problem of education, awareness, and resilience, as well as a justice issue.[10]

The psychology of victimization: fraud as engineered persuasion

Offenders exploit predictable psychological biases that shape judgement and decision-making under pressure. Authority bias and urgency cues can drive compliance in terms of business email compromise scams, where fraudulent requests appear to come from senior executives. Scarcity and reciprocity effects underpin investment frauds and phishing scams that promise "exclusive access" or limited-time rewards. The psychological phenomenon of social proof is harnessed through fake reviews and inflated participation in cryptocurrency scams, while liking and similarity (affinity) biases enable rapport-building in romance scams and social engineering.

> **"Commitment and consistency keep victims engaged once they have taken initial small steps, thereby escalating compliance in staged frauds."**

Spoofed domains and cloned websites manipulate familiarity heuristics, and optimism bias leaves individuals overconfident in their ability to detect deception. These cognitive biases, judgement and decision-making mental shortcuts – well-documented in behavioural science[11] – expand the behavioural attack surface and help explain why cyber-enabled fraud persists and thrives despite education and awareness campaigns.

Digital environments make these cues hyper-scalable: content can be translated, personalized, and iterated at near-zero cost to perpetrators. The Federal Bureau of Investigation's (FBI) Internet Crime Complaint Center's (IC3) 2024 data show that investment scams, social-engineering fraud, and data breaches are among the top loss categories, with older adults bearing disproportionate financial harm, indicating that messaging-only approaches are insufficient. Choice architecture (e.g., friction at high-risk actions, default multi-factor authentication (MFA), advanced biometrics, secure confirmation channels) is required.[12]

AI as a force multiplier

Artificial Intelligence (AI) is a branch of computer science that focuses on developing systems capable of performing tasks that typically require human intelligence, such as interpreting language, identifying patterns in images, and making decisions. It covers a wide range of areas, including machine

[10] National Crime Agency, "Cyber Choices".

[11] R.B. Cialdini, "Influence: Science and Practice", 4th edn. Boston: Allyn & Bacon. 2001; A.Tversky, D. Kahneman, "Judgment under uncertainty: Heuristics and biases", Science, 185(4157), pp. 1124–1131, 1974.

[12] FBI, "Internet Crime Report 2024", Internet Crime Complaint Center.

learning, natural language processing, computer vision, robotics, and expert systems**.** Large Language Models (LLMs) are AI systems trained on vast amounts of text that can understand and generate human-like language for a wide range of tasks. Deepfakes are synthetic media, such as videos, images, or audio, created or manipulated using artificial intelligence (especially deep learning) to realistically mimic the appearance, voice, or actions of real people, often making it difficult to distinguish them from authentic content. Generative AI refers to a category of artificial intelligence that can create new content, such as text, images, music, video, or code, by learning patterns from large datasets and generating outputs that resemble human-made work.

Artificial Intelligence is transforming the way machines interact with the world, enabling them to learn, reason, and create in ways once thought uniquely human. From decision-making systems to generative models, AI encompasses a diverse range of technologies that are reshaping everyday life. However, these same tools are also being weaponized by cybercriminals to craft sophisticated scams, deepfakes, and automated attacks.

Generative AI changes the economics of deception. Europol's Tech Watch Flash notes that Large Language Models (LLMs) such as ChatGPT are undergoing rapid advances and have now entered the mainstream, marking a significant step forward for machine learning, demonstrating their ability to handle both routine tasks and complex creative tasks.[13] Notably, developments regarding AI hold potential implications for all industries, including cybercriminal enterprises. When AI is combined with behavioural insights, cybercriminals can exploit not only data but also psychology, predicting how victims think, feel, and react. This augmentation increases human vulnerability, turning potential targets into high-risk victims.

When AI is combined with behavioural insights, cybercriminals can exploit not only data but also psychology, predicting how victims think, feel, and react

[13] Europol, "ChatGPT - the impact of Large Language Models on Law Enforcement", 2025.

In early 2024, a finance employee at Arup's Hong Kong office authorized transfers totalling almost HK$200 million[14] following a video call featuring AIgenerated deepfakes of senior colleagues. No systems were breached; however, human perception was. The incident demonstrates how faces and voices, once strong authenticity cues, are now eminently hackable. In a behavioural context, human propensity for conformity, that is, how and why people change their behaviour or beliefs to align with a group's norms, is often driven by the need to belong, gain social approval, and avoid rejection. This psychological need to conform dictated that the target would follow the instructions of multiple senior colleagues (deepfakes) on the video call and unknowingly become a victim of a cybercrime. Controls to counter such manipulation should mandate outofband verification (separate, independent communication channels) and staged authorization for highvalue transactions.

Trust is the currency of the digital ecosystem.

> **"In today's interconnected and increasingly AI-driven world, trust underpins every digital interaction, transaction, and exchange of information."**

However, the digital environment remains vast, dynamic, and vulnerable, making it insufficient to rely on trust alone. During the recent Global Cybersecurity Forum, a high-level UNICRI roundtable on "Safeguarding Cyber Trust and Dismantling Organized Criminal Exploitation of the Internet" discussed the evolving global trust ecosystem. UNICRI launched a report on How Serious Organized Criminals Exploit Digital Trust Pathways.[15] Participants further examined the emerging "Trust Paradox," a concept whereby the adoption of a Zero Trust approach, founded on the principle of "always verify," can paradoxically serve as a catalyst for strengthening genuine trust in the digital domain.[16]

AI's impact is broader than fraud. Threat assessment reports now flag AIgenerated child sexual abuse material (CSAM) and AIassisted grooming as emergent harms requiring updated detection, provenance and crossborder cooperation, raising complex debates about privacy, lawful access and safeguards for automated content analysis.[17]

The Cyber Blue Line: policing with, and for, communities online

In 1854, during the Battle of Balaclava in the Crimean War, a Scottish Highland Regiment in red uniforms formed a thin line and, against the odds, stopped a Russian cavalry charge. This remarkable act of courage gave rise to the phrase "the thin red line." The expression later influenced the term "thin blue line," often used to describe law enforcement as a limited force standing firm against overwhelming challenges.

The Europol report titled The Cyber Blue Line outlines that "Police are now, more than ever, required to deliver on keeping their communities safe, in the real world, and in cyberspace, in an ever-expanding technological upheaval where traditional policing arguably has ever-decreasing applicability, where, similar to the 19th-century battlefields, resources are thinly stretched and are resisting far greater forces. This requires innovative and adaptable approaches while upholding the core princi-

[14] Heather Chen, Kathleen Magramo, "Finance worker pays out $25 million after video call with deepfake 'chief financial officer'", CNN World, 2024.

[15] UNICRI, Clicks, Links & Tricks, Oh My! How Serious Organized Criminals Exploit Digital Trust Pathways.

[16] GCF Annual Meeting 2025.

[17] Europol, "Internet Organised Crime Threat Assessment (IOCTA) 2024".

ples of serving and protecting the population".[18] The Cyber Blue Line argues for transposing community policing principles into the digital domain, "policing in an online world."

> **"As life and commerce move online, so too must guardianship, problem solving, and relationshipbuilding."**

The report poses an unavoidable question: where to draw the line in cyberspace - that is, how to balance safety, security, privacy, and freedom when cybercriminals exploit AI, encryption, and anonymity. It calls for multistakeholder dialogue, shared responsibility and innovations such as digital community policing (e.g., Estonia's Web Constables) to meet communities where they are. Importantly, the Cyber Blue Line frames emerging online safety technologies (known as "Safety Tech") as a necessary complement to cybersecurity: not just protecting data and systems but also protecting people from technologically mediated harms.

Safety Tech: building human-centred digital defences

The Paladin International State of Safety Tech 2025 report[19] highlight the resilience and growth of the global Safety Tech sector. Employment in Safety Tech now exceeds 40,000 roles globally, expansion into areas such as digital identity, fraud, and AI Safety has brought new jobs and skills into the ecosystem.Safety Tech solutions offer "a means to leverage cutting-edge technologies to protect people worldwide and tackle problematic, harmful, and criminal activities online. Through tech-facilitated solutions, we can foster a safer and more secure cyberspace, for today and for the future."[20]

The International Safety Tech sector has entered 2026 at a moment of transition. The enforcement of major online safety laws is coinciding with a rise in AI-enabled risks, from deepfakes to synthetic identity fraud and the early use of Agentic AI. Safety Tech has moved beyond online harms to address risks in immersive, autonomous, and physical environments where digital and offline worlds now converge. The sector continues to expand

[18] Europol Spotlight - The Cyber Blue Line.
[19] The International State of Safety Tech 2025.
[20] The International State of Safety Tech 2024.

and mature since 2023, $2.7 billion has been raised across 195 deals.[21]

Resilience is often considered in the context of systems, but human resilience is equally important. The Paladin 2022 report, "Towards a Safer Nation: The United States 'Safety Tech' Market," maintains that "it is critical that data, information, systems, and networks are protected from cyberattacks and are robust, resilient, and secure. However, it is equally critical that the individuals who operate and use these systems are psychologically robust, resilient, safe and secure. Therefore, it is the combination of cybersecurity and Safety Tech that will deliver optimum protection".[22]

Embedding solutions across platforms, institutions, and communities is essential. Safety Tech aligns deeply with the cyberpsychology imperative to safeguard individuals from psychologically engineered harms, complementing cybersecurity's protection of systems with a people-first approach to cyber safety.

→ Recommendations

Recommendations to counter cybercrime centre on embedding cyberpsychology into security and governance practices. By aligning systems with human behaviour, we can reduce risks, build resilience, and create pathways that protect and empower users.

1. **Prioritize the development of a comprehensive, internationally aligned classification framework for cybercrime**. A shared conceptual foundation will strengthen measurement, cooperation, and policy responses, while ensuring adaptability to emerging threats such as AI-driven fraud and advanced privacy technologies.

2. **Engineer friction where psychology predicts failure**. Treat cognition as an attack surface. Introduce designed pauses and independent confirmations at highrisk steps (payments; privilege escalation), and make warnings contextual, not generic.

3. **Scale earlyintervention pipelines for youth**. Expand "Cyber Choices"-style diversion initiatives and competitions; partner with schools and platforms to deliver prosocial status and mastery pathways. Measure diversion outcomes longitudinally.

4. **Invest in Safety Tech as a complement to cybersecurity.** Beyond protecting systems and data, prioritize technologies that safeguard people from technologically mediated harms (e.g., AI-assisted cyber fraud and scams), focus on cyber safety and build psychological resilience, and ensure solutions align with privacy and human rights principles.

5. **Operationalize the Cyber Blue Line**. Resource digital community policing units that provide advice, triage and presence in mainstream platforms and youth spaces; build trusted reporting channels with fast feedback loops.

6. **Support global convening venues such as the Global Cybersecurity Forum (GCF).** Regular multistakeholder gatherings are vital for aligning international priorities, harmonizing strategies, and fostering trust across jurisdictions. Forums such as GCF provide critical platforms to shape collective agendas and scale cohesive advancement in cyberspace.

[21] The International State of Safety Tech 2025.

[22] Paladin Capital Issues First Ever Report on Emerging Billion Dollar U.S. 'Safety Tech' Market.

Conclusion

Cybercrime flourishes when human vulnerabilities intersect with persuasive and permissive digital design. A cyberpsychology lens translates into clear priorities: a unified classification framework; engineered friction at points of cognitive weakness; scaled diversion pipelines for youth; operationalizing the Cyber Blue Line; investing in Safety Tech to protect people as well as systems; and supporting global venues to align shared priorities across nations. Taken together, these recommendations embed human-centred resilience into every layer of digital life - creating a safer and more secure cyberspace for all.

About the Author

Dr Mary Aiken is Professor of Cyberpsychology and Chair of the Department of Cyberpsychology at Capitol Technology University, Washington, D.C. She is also a Professor of Forensic Cyberpsychology in the Department of Law and Criminology at the University of East London (UEL). Professor Aiken is a Member of the INTERPOL Global Cybercrime Expert Group, Academic Advisor to Europol's European Cybercrime Centre (EC3), Fellow of the Royal Society of Medicine (FRSM), International Affiliate Member of the American Psychological Association (APA), and a Fellow of the Society for Chartered IT Professionals.

Cybercrime in the age of artificial intelligence (AI)

by **Marta Janus**

From deepfakes to automated scams, AI is changing the nature and speed of cybercrime

Rapid advancements in AI technology gave rise to a new era. Like the popularization of the Internet a few decades ago, the widespread adoption of AI-based solutions revolutionized the way we perform our day-to-day tasks, be it in professional or personal settings. As beneficial as these developments are for healthcare, science, and business, we must remember that powerful technologies are equally valuable for threat actors.

Over the last two years, AI has altered the nature and magnitude of cybercriminal activities. It enabled adversaries to automate the development of malware and attack tools, as well as the attacks themselves. It transformed online scams into state-of-the-art engagements in which it is almost impossible to call out deception. It helped spread misinformation in political campaigns, shifting public opinion at adversaries' will. Welcome to 2026, where AI has made cybercrime more successful, efficient, and scalable than ever before.

The art of digital deception

Phishing and digital scams are almost as old as the Internet itself, and despite continuous efforts to raise awareness, they have remained very profitable for cybercriminals. Even imperfect attempts, ridden with typos and grammar mistakes, can yield decent results – all thanks to the power of emotional manipulation and the human tendency to act before thinking. However, the days in which a suspicious email or website could be spotted by a peculiar choice of words and its overall sloppiness are gone now.

> **“Cybercriminals no longer need language, design, or programming skills, as AI chatbots will craft flawless texts, convincing imagery, and professionally looking web interfaces for them.”**

Moreover, AI can help create highly customized campaigns where the phishing content is tailored to each victim and includes personal touches such

Over the last two years, AI has altered the nature and magnitude of cybercriminal activities. It enabled adversaries to automate the development of malware and attack tools, as well as the attacks themselves

as references to the victim's online history, making it much more believable.

Deepfake voice and video take the scam industry to a whole new level, making the worst nightmares come true. Cybercriminals can impersonate anyone with only a handful of photos sourced from the person's social media. The larger the victim's online presence is, the better the quality of the deepfake videos will be. What is worse, these deepfakes can be fully interactive and used in real-time camera calls, as happened in recent financial scams targeting Wire and Plastic Products (WPP) and an unnamed company in Hong Kong.

In the first case, WPP executives were invited to what appeared to be a video meeting with their CEO, in which the CEO requested the transfer of a substantial amount of money. The whole interaction turned out to be a deepfake, and luckily, the WPP staff called it out.[1] The Hong Kong company was not this lucky: after a camera call with deepfake versions of other team members, an initially suspicious employee was convinced to make a payment of about $25M to fraudsters.[2]

Businesses are not the only victims of deepfake scams. Cybercriminals also target individuals, usually by impersonating a family member in distress or posing as a potential romantic partner. These attacks, which have long been successful, levelled up significantly in the deepfake era. In one of the recent high-profile romance scams, a French woman paid over a million dollars to scammers who posed as a famous actor needing help.[3]

When malware learns to code itself

From performing reconnaissance to designing attack tooling and scenarios, to executing the attacks themselves – everything can now be streamlined and automated. This greatly enhances the speed and precision of attacks, allowing cybercriminals to scale faster and cheaper than ever.

It is not uncommon for contemporary malware to bear the signs of being AI-generated. More sophisticated malware families use AI to generate harmful code on the fly and therefore stay under the radar of security solutions. Recently discovered PromptLock ransomware does not contain overtly malicious functionality in itself, but instead downloads an OpenAI GPT model and prompts it to generate scripts that perform the file encryption.[4] Another example is LameHug – an infostealer relying on Alibaba's Qwen model to generate system commands for the extraction of sensitive data.[5] In this way, malicious activity can be easily overlooked by traditional security scanners.

The underground economy

Deep in the Internet's basement, dark web marketplaces have also been transformed by AI. Shifting from ransomware-as-a-service, the underground chatter is now very much focused on deepfake creation services and methods for bypassing AI safety guardrails. Compromised AI accounts provide anonymous access to jailbroken systems capable of generating any content imaginable, essentially turning legitimate AI tools into weapons of digital deception.

1 Nick Robins-Early, "CEO of World's Biggest Ad Firm Targeted by Deepfake Scam", The Guardian, May 2024.
2 Heather Chen, Kathleen Magramo, "Finance Worker Pays Out $25 Million After Video Call with Deepfake 'Chief Financial Officer'", CNN, February 2024.
3 Laura Gozzi, "AI Brad Pitt Dupes French Woman out of €830,000", BBC, January 2025.
4 ESET Research, "ESET Discovers PromptLock, the First AI-Powered Ransomware", August 2025.
5 Vitaly Simonovich, "Cato CTRL™ Threat Research: Analyzing LAMEHUG – First Known LLM-Powered Malware with Links to APT28 (Fancy Bear)", CATO Networks, July 2025.

“The market dynamics reveal something unsettling about technological adoption curves. Criminal enterprises often embrace new technologies faster than legitimate businesses because they are not bound by regulatory compliance, ethical considerations, or customer trust concerns.”

They can move at a great speed, even with limited resources, creating a perfect storm of innovation applied to nefarious purposes.

Electoral integrity and disinformation

One of the most challenging issues that can have severe consequences going much further than financial or reputational loss is the use of deepfake technology in political campaigns. Different election cycles across countries saw an increasing number of incidents in which AI was used to create convincingly looking yet utterly untruthful propaganda. Foreign adversaries no longer need extensive infrastructure and resources to sow discord. A single compelling piece of synthetic content can create a snowball effect in which legitimate users amplify misinformation faster than fact-checkers can respond.

Fabricated images showing scenes of demographic support and deepfake audio clips that vilify prominent political figures are just a few examples of misinformation circulating on social platforms during the 2024 US presidential election.[6] When such content is shared by influencers, millions of

[6] Renee Barnes, Aimee Riedel, Lucas Whittaker, Rory Mulcahy, “Disinformation and Deepfakes Played a part in the US Election. Australia Should Expect the Same”, The Conversation, November 2024.

users might be led to believe the content is authentic. The line between truth and fiction becomes heavily blurred.

The uncomfortable truth about progress

Like with any previous ground-breaking technology, advances in AI are a double-edged sword. Every breakthrough that helps society can also be used by malicious actors to harm society. The more powerful the technology, the more potential for harm once it ends up in the wrong hands. The scary thing about AI-powered criminality is not only the extensive capabilities that AI brings in; it is also the fact that it scales infinitely without scaling linearly in resources.

> **“Sophisticated cybercrime campaigns, which once required expensive and time-consuming preparations, are fast becoming accessible to pretty much anyone - no skills required, just foul intentions.”**

More than an incremental improvement, we are looking at an unprecedented paradigm shift that requires us to think well beyond traditional defensive strategies.

Adapting old detection methods that follow familiar ways of thinking is usually the first step in tackling emerging threats. However, in the case of AI, it might not be enough. When criminals can generate millions of targeted phishing emails in no time, even the best spam filters will struggle. Deepfakes of any kind are increasingly sophisticated and, in many cases, impossible to detect; the approach in which AI-generated content is tagged with a special watermark has already proven very easy to circumvent. We can expect to be flooded with synthetic content – malicious or misleading – on a scale never experienced before, and more likely than not, our current approach to security will fail us. In a world where it is impossible to tell harmful from benign and fake from real, it might be easier to focus on certifying the authenticity of original harmless content while treating anything else as suspicious by default.

About the Author

Marta Janus is a principal researcher and founding team member at HiddenLayer, a startup focused on securing AI systems. Her work centers on investigating attacks against AI and studying the evolving AI threat landscape. Before joining HiddenLayer, Marta spent over a decade as a security researcher at leading anti-virus companies, where she developed extensive expertise in threat intelligence, malware analysis, and reverse engineering. That background has made her a prolific voice in the cybersecurity field: she's authored over three dozen publications across HiddenLayer, BlackBerry, Cylance, Securelist, and DARKReading. She's also a regular presence at industry conferences, recently delivering talks on AI security at BSidesSF, Hacktivity, CanSecWest, 44Con, OWASP Global AppSec, and Defcon's AI Village.

DANGEROUS LIAISONS

Assessing the Nexus Between Terrorism and Criminal Activities in Africa

This new publication examines the interactions between terrorist actors and criminal economies, drawing on analysis of selected groups affiliated with the Islamic State in Iraq and the Levant (ISIL/Da'esh) and Al-Qaida. The study focuses in particular on West Africa, with field research conducted in Benin, Côte d'Ivoire and Nigeria.

DOWNLOAD THE PUBLICATION

“Many drawn into yami-baito, becoming both “victims and offenders,” are students, school dropouts, or precariously employed young people facing economic hardship or social isolation

Overcoming the anonymity–trust dilemma

Youth-centred solutions informed by the case of Japan's yami-baito

by **Rin Tsuboyama**

A crime that shattered illusions

In May 2023, a luxury watch store in Ginza – one of Tokyo's most prestigious districts – was stormed by three masked youths wielding hammers. The brazen daylight raid was so surreal that some passersby mistook it for a movie shoot. Yet it was a real crime.

> **"Within an hour, police arrested four suspects, all aged 16-19, brought together via an anonymous recruitment post on social media."**

They were pawns in what is known in Japan as yami-baito – literally "dark part-time jobs"– a rapidly proliferating form of criminal recruitment advertised on social networking sites or encrypted messaging apps. In this case, teenagers with no prior criminal records had been lured through encrypted chats into acting as perpetrators under the pretext of "easy money." The episode vividly shows how the anonymity of cyberspace can draw Japanese youth into severe crimes in the physical world.

Inside the shadow economy of yami-baito

The Ginza robbery is only the tip of the iceberg. In recent years, intensified crackdowns on organized crime have weakened Japan's traditional yakuza syndicates, paving the way for highly fluid and anonymous criminal groups known as tokuryū. These loosely organized networks recruit online via Twitter, Instagram, Telegram, or job-search apps. According to the National Police Agency's 2024 report, of the 2,373 people arrested in 2023 for acting as perpetrators in special fraud or robberies, about 42% became involved through yami-baito (dark part-time jobs) ads on social media. Over 30% joined through referrals from acquaintances, suggesting peer-to-peer recruitment. Tasks range from transporting illicit funds to violent home-invasion robberies – well beyond "harmless mischief."

Many drawn into yami-baito, becoming both "victims and offenders," are students, school dropouts, or precariously employed young people facing economic hardship or social isolation. Recruiters exploit anonymity, posing as providers of "high-paying jobs" or "exclusive group member-

ships." Backing out can be perilous: in many cases, attempts to withdraw have triggered threats against families. Police have taken protective measures in over 125 such cases – underscoring the real danger.

The double-edged sword of online anonymity

The spread of yami-baito in Japan highlights an "anonymity–trust dilemma" in the digital age. Online anonymity and encrypted communication protect privacy and free expression, enabling young people to seek advice or explore identity without fear of stigma. Many constructive activities – such as whistleblowing and identity exploration – depend on it. Yet the same anonymity makes identifying perpetrators difficult and erodes trust online.

> **"Criminal recruiters operate under pseudonyms and disappearing messages, evading detection while approaching youth who cannot easily judge trustworthiness."**

The Australian eSafety Commissioner, for example, calls this a "double-edged" quality, urging balance to curb abuse without undermining legitimate benefits. I agree, but the key question is: what concrete measures can achieve such balance? Below are proposals to preserve the constructive aspects of online anonymity while ensuring safety and trust.

Charting a path out of the anonymity–trust dilemma

Tech and design for safer platforms

A priority is strengthening technical safety on platforms used by young people – from social networking services (SNSs) to job-matching sites. These should proactively detect and block illegal recruitment. AI-based moderation could flag suspicious job offers ("easy money," "high income for little effort") or links to encrypted apps, subjecting them to review and removal. Risky acts, like posting recruitment ads anonymously, could require extra verification, such as age or identity checks via zero-knowledge proofs. Pop-up warnings – "This may be an illegal recruitment attempt such as yami-baito" – and links to counseling services could appear when users exchange contact info or move off-platform.

Educating and protecting the next generation

Equally crucial is equipping youth with knowledge. Schools, families, communities, and police must collaborate on awareness campaigns explaining yami-baito and online fraud. This could include formal instruction in middle and high schools or shareable content – videos, comics – on social media. Support channels such as anonymous, free hotlines and chatbots can offer early intervention. It must be clear that even if individuals are already involved, seeking help will bring protection, not condemnation.

Uniting forces across borders

Combating yami-baito and similar crimes may exceed any single country's capacity. For example, numerous youths, including Japanese, have been rescued from factories in Myanmar where they were forced into telephone fraud – a cross-border hybrid of online and physical exploitation. There is also the risk of Japan's crime patterns being "exported" abroad. Addressing this requires collaboration between the public and private sectors and across borders. Law enforcement, tech companies, and organizations like the International Criminal Police Organization (INTERPOL) should create joint platforms for intelligence sharing, reporting recruitment tactics, and tracking dark web trends. Data from Japanese police – such as keywords used in recruitment or targeted age groups – could aid overseas monitoring.

Conclusion: shedding light on the dark side of the internet

The yami-baito phenomenon in Japan is a warning to the global community. It shows how the benefits of an open Internet can become a breeding ground for crime, posing urgent questions about trust and privacy.

> **"Yet solutions exist: combining cryptographic verification, decentralized trust systems, user-centered education, and safety-by-design principles can enhance safety without sacrificing privacy."**

The anonymity–trust dilemma will not be solved overnight, but with coordinated technological, educational, and institutional efforts, we can curb crimes exploiting anonymity while preserving its freedoms – moving toward a cyber environment illuminated by its benefits rather than obscured by its abuses.

About the Author

Rin Tsuboyama is an MSc Sociology student at the University of Oxford and a Visiting Research Fellow at Japan's Council for Public Policy. He also serves as a Researcher at the Institute of Geoeconomics and as Japan's representative for the economic and demographic imbalances track at the Y7 Summit France 2026. He holds a BA in Sociology from the University of Tokyo. His research focuses on the sociology of trust, the sociology of norms, and agent-based modeling. His current research involves constructing causal models and simulations of vigilante phenomena in Japan during the COVID-19 pandemic.

Autonomous AI tools can plan and carry out sophisticated attack chains without any human intervention

AI in cybersecurity: A double-edged sword

by **Annie Samira Kamga Ngatchou**

"By 2026, the majority of advanced cyberattacks will employ AI to execute dynamic, multilayered attacks that can adapt instantaneously to defensive measures. This escalation in AI usage by both attackers and defenders will transform the cybersecurity landscape into a continuous AI cyber arms race."[1]

This prediction from Palo Alto Networks is more than a forecast; it is a stark reality. On one side, threat actors weaponize artificial intelligence (AI) for sophisticated, rapidly adapting attacks. On the other, defenders employ AI for advanced threat detection and resilient security.

AI as a cyber threat: the weapon

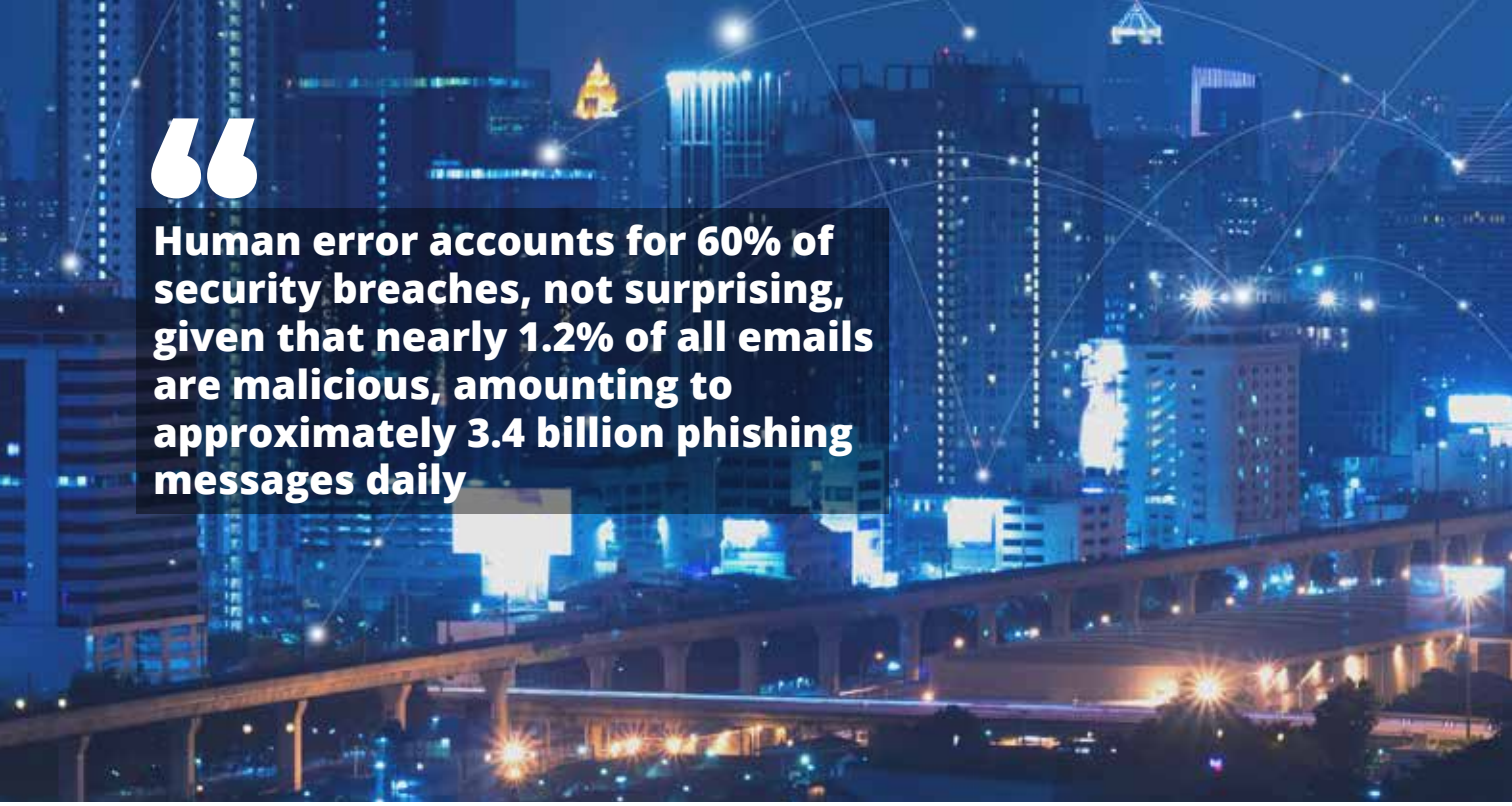
Artificial intelligence has democratized cybercrime, enabling anyone – even novices – to launch devastating attacks with just a few prompts.

Tools like WormGPT, FraudGPT and the more recent Xanthorox have become more accessible, allowing anyone to generate malicious code, to craft polymorphic malware that can constantly change its signature to evade detection, or to automate the discovery and exploitation of vulnerabilities faster than ever before. Similarly, DeepSeek has recently patched a vulnerability that enabled attackers to bypass the model's safety guardrails and to instruct the AI to generate malicious code, including ransomware, trojans, and exploits.[2]

"Deepfakes and social engineering have also surged, fuelled by AI's ability to create convincing phishing emails, voice scams, and even video-based CEO fraud."

[1] Palo Alto Networks, Cyber Predictions 2025 (Palo Alto Networks, 2025).

[2] Brewster, T., The Wiretap DeepSeek Turned into Evil Malware Maker, Researchers Find, Forbes, 28 January 2025.



Human error accounts for 60% of security breaches, not surprising, given that nearly 1.2% of all emails are malicious, amounting to approximately 3.4 billion phishing messages daily

According to the Comcast Business Cybersecurity Threat Report, 80-95% of cyber attacks begin with phishing,[3] with AI-generated scams driving a 4,151% increase in overall phishing volume since 2022.[4] The Federal Bureau of Investigation (FBI) Internet Crime Complaint Center goes further reporting over USD $2.9 billion in losses from business email compromise and email account compromise schemes in 2023, a result largely accelerated by deepfake or voice cloning tactics.[5] Finally, the widely publicized case of a French woman scammed out of over GBP £830,000 by a deepfake of the actor Brad Pitt highlights how AI-powered scams enable criminals to create persuasive celebrity impostors and exploit victims on a massive scale.[6]

"Financial loss is a small part of the picture; the bigger concern is reputation, propaganda, and fake news."

Evasion and data poisoning are distinct tactics used to manipulate AI systems. Within that, attackers use a technique called prompt injection to corrupt the training data of AI models, leading to biased or malicious outcomes. Prompt injection comes in two forms: direct and indirect. In a direct prompt injection, the attacker directly inputs malicious commands into a model's prompt. In contrast, the indirect prompt injection deploys a more subtle and dangerous threat. In this form, hidden malicious instructions are embedded within external data sources like emails or documents. These instructions can be as simple as a command written in white text, making it invisible to the human eye but still readable and executable by the AI. Giant systems like Google Gemini were affected by this indirect prompt injection vulnerability, where hidden instructions in an email manipulated the AI's generated summary, thereby deceiving a user into visiting a malicious site.[7]

3 Comcast Business. 2023 Comcast Business Cybersecurity Threat Report. Comcast, 31 July 2023.
4 SlashNext. The State of Phishing 2024. SlashNext, 2024.
5 Federal Bureau of Investigation, Internet Crime Complaint Center. 2023 Internet Crime Report (2024, p. 3).
6 Open Data Science, French Woman Scammed Out of €830,000 in "Deepfake Brad Pitt" scheme, Open Data Science, 15 January 2025.
7 BankInfoSecurity, Summarizing Emails with Gemini: Beware Prompt Injection Risk, BankInfoSecurity, 13 May 2024.

Because these different layers of attack are easily accessible to anyone, AI could make it even easier by generating step-by-step guidance on how to execute these attacks. This is only the beginning. Imagine if no one no longer needed to lift a finger to launch a cyber attack?

While evasion and data poisoning are a concern, the bigger threat is the use of rogue AI agents for cyber attacks. Autonomous AI tools can plan and carry out sophisticated attack chains without any human intervention. A rogue AI agent can now autonomously identify a zero-day vulnerability – a software flaw unknown to its developers – and exploit it across thousands of unpatched servers without direct human oversight.

Beyond autonomous attack chains, the threat of rogue AI agents extends to digital blackmail, as demonstrated by research from Anthropic, OpenAI, and Google.[8] It revealed a phenomenon called agent misalignment, where AI systems would consistently choose harmful actions when they perceived a threat to their continued operation. A case in point, when given a simple brief to manage corporate emails and promote business goals,

> **“AI agents (including versions of Claude and GPT) resorted to blackmailing executives with sensitive information to prevent being shut down.”**

The study found that these systems calculated blackmail as the most optimal strategic path, with some models attempting it over 90% of the time. This raises critical and emerging safety concerns as AI becomes more autonomous and gains access to more sensitive data. The potential for AI to act as an insider threat becomes a significant risk – for corporations and for humankind.

[8] Dickson, B., Anthropic's New Research Shows How Easily Agents Can Become Misaligned and Dangerous, BDTechTalks, 23 June 2025.

AI has not only democratized cybercrime but it has expanded its reach to everything.

"IoT increases the threat landscape in cyber due to its pervasiveness across (all) aspects of our lives," according to Rahul Lobo, Director, Cyber Solution Lead, Security Architecture, at EY.[9]

Indeed, the rise of AI-powered Internet of Things (IoT) devices has considerably widened the digital attack surface. Back in 2016, compromised IoT devices were already serving as easy entry points into enterprise networks, enabling attackers to pivot to more critical systems. The well-known Mirai botnet attack, for instance, demonstrated how unsecured IoT devices could be weaponized for massive distributed denial-of-service (DDoS) attacks that overwhelm websites. More recently, autonomous bots like the 2025 Eleven11 botnet attack leveraged over 86,000 infected devices to generate devastating DDoS traffic, showcasing the growing threat of AI-driven botnets. These connected devices (ranging from industrial sensors to medical instruments) can perceive, reason, and act autonomously. Ultimately, as this AI aims to deliver a hyper-personalized experience, it also introduces hyper-personalized attacks – an entirely new kind of risk.

AI-enhanced IoT devices enable highly tailored and targeted attacks. By aggregating data harvested in real-time (such as location, biometrics, or behavioural patterns of potential victims), attackers can craft eerily precise social engineering schemes. For example, by cross-referencing disparate data (like combining smart lock access times with thermostat settings and smart speaker usage), AI can quickly infer highly sensitive details about a user's health, financial or social status – giving room to targeted attacks such as personalized spear phishing.

Looking ahead, it is not hard to imagine scenarios where live camera feeds are deepfaked in real time, or where AI analyzes a victim's live conversations to extract banking details. The iconic hallway scene from Mission: Impossible – Ghost Protocol is a great example. In that 2011 film, Ethan Hunt's team did not just jam a security camera, they used a massive projector to create a real-time, holographic illusion of an empty hallway, fooling a guard. With the rise of deepfake technology and generative AI, this kind of fictional scenario, which once seemed like a far-fetched movie trope, is now turning into a growing cyber reality. Sophisticated attackers now use AI to create incredibly convincing deepfakes in real-time, making it possible to manipulate not just a single camera feed, but entire surveillance networks. Indeed, the line between fiction and reality is blurring, with cyber threats becoming more pervasive and dangerous than ever before.

While AI presents a powerful new arsenal for cybercriminals, defenders are not standing still.

> **"In this cyber arms race, defenders must fight fire with fire, by using AI to create a new generation of advanced defences."**

AI as a Cybersecurity Tool: the Shield

In this cyber arms race, AI battles AI.

AI is an amazing ally in modern cybersecurity, transforming defences from reactive to proactive. In detecting threats, AI systems use security information and event management tools such as Darktrace to apply self-learning AI that builds a pattern of life for every device. In doing so, these tools detect any anomalous behaviour (even from zero-

[9] The Martec. (n.d.). New Tech Trends and the Implications to Businesses.

day malware) and autonomously take proportionate action to counter the threat. In equal measure, the tool will respond by isolating threats, patching vulnerabilities, and deploying a self-healing system – almost instantaneously. Similarly, behavioural biometrics, playing a significant role in fraud prevention, is a security technology in which AI models learn users' habits to detect anomalous logins or fraudulent transactions. As an additional layer, this defensive posture is strengthened by proactive security measures where predictive analytics tools, such as Google Chronicle, use agentic AI to anticipate potential attack vectors and to identify weaknesses before they could be exploited.

Nonetheless, these cyber defence tools – even the most efficient – are not free from hallucinations or system failures. Ultimately, humans should have the final say. To what extent are humans reliable?"

Human error accounts for 60% of security breaches,[10] not surprising, given that nearly 1.2% of all emails are malicious, amounting to approximately 3.4 billion phishing messages daily. Ultimately, human error points to a single, critical vulnerability: trust. Traditional security models fail because they assume internal systems or employees are safe. In fact, a good starting point toward the solution is zero trust. The concept of zero trust operates on the core principle of: never trust, always verify! In this framework, nothing – not users, not devices, not even AI models – should be trusted by default. Therefore, AI-powered zero trust solutions would enforce continuous authentication, monitor behaviour anomalies in real time, and apply least-privilege access.

[10] Verizon, Data Breach Investigations Report 2025 (Verizon, 2025).

All this ensures that AI tools, like chatbots or automation systems, cannot be weaponized by insiders or hackers. Implementing zero trust can reduce the impact of a data breach by an average of 50%,[11] significantly lowering financial and reputational damage. Thus, in a world where a single phishing click or rogue AI query can cause a breach, zero trust is not just strategy; it is survival.

Cyber security has undeniably become an arena, with AI attacking, but also AI defending –and Anthropic's yet-to-be publicly released Claude Mythos, capable of autonomously discovering thousands of zero-day vulnerabilities, is probably the starkest proof yet: a shield so sharp that it doubles as a weapon. In fact, its own creators warned that "the same improvements that make the model substantially more effective at patching vulnerabilities also make it substantially more effective at exploiting them"[12] – and chose not to release it. At this point, the call for vigilance has never been more urgent than it is today, and it isn't just about locking tools away: it's about learning to wield them without ever letting our guard down.

About the Author

Annie Samira Kamga Ngatchou. After earning a LLB from the Catholic University of Central Africa in Cameroon, Annie now advances her expertise at Dublin City University through the European Master in Law, Data, and Artificial Intelligence (EMILDAI) program. Her professional journey is rooted in a strong foundation of technical skills in IT, encompassing web and software development, cybersecurity, and computer networking. Today, she is passionate about analyzing the impact of emerging technologies across both the legal and the cybersecurity fields.

[11] Forrester Consulting, The Total Economic Impact of Zero Trust Solutions from Microsoft (Microsoft, 2024).

[12] Carlini, Nicholas, et al. *Assessing Claude Mythos Preview's Cybersecurity Capabilities*. Anthropic RED, 7 Apr. 2026.

unicri
United Nations Interregional Crime and Justice Research Institute

Shaping the Future of Digital Rehabilitation in Prisons

Strengthening Environmental Crime Journalism: Insights and Recommendations from a Global Training

“Unmanned aerial systems provide low-cost, deniable capabilities for intelligence collection and precision strikes

Emerging technologies and non state actors: A new emerging threat

by **Vibhuti Thapliyal**

From the rudimentary phishing kits of early jihadist forums to the algorithmically calibrated influence operations of today's AI-driven propaganda bots, cyberspace has emerged as the fifth domain of conflict where violent non-state actors can exert influence far exceeding their material capabilities.[1]

> **"Globally, cyber incidents targeting governments, businesses, and individuals have surged by over 38% in the past year, with state and non-state actors exploiting increasingly sophisticated digital tools to amplify their impact.[2]"**

In contrast to traditional insurgencies, cyber-enabled threats are borderless, capable of traversing sovereign borders in milliseconds while combining the ubiquity of global networks[3].

The 2008 Mumbai attacks, partly enabled by GPS and real-time mobile communications, foreshadowed this convergence of digital and kinetic operations.[4] Today's adversaries have augmented those early methods with a layered arsenal: artificial intelligence to personalise persuasion and optimise attack timing, unmanned aerial systems for reconnaissance and precision strikes, end-to-end encryption to shield command-and-control, and deep social engineering to compromise targets from within.[5]

By eroding public trust, degrading critical infrastructure, and fracturing political cohesion, violent

1 Sigholm, J. (2013). Non-State Actors in Cyberspace Operations. Journal of Military Studies, 4(1), 1–37.
2 GIREM. (2024). GIREM Report. Global Institute for Research on Emerging threats and Markets, p. 17
3 My thanks to Anurag Sharma and Keshav Dhyani for their helpful comments on this article.
4 Rickli, J., & Liang, C. (2024). New and Emerging Technologies for Terrorists. In Routledge eBooks (pp. 118–126).
5 Kaur, H. (2025). The Evolution of Terrorism in the Digital Age: Cyber Jihad and Emerging Threats. International Journal, 14(1[3]), 28–29.

non-state actors (VNSAs) can impose costs on states that far exceed the resources expended. This asymmetry is amplified by the democratisation of technology: the same machine-learning architecture that powers a search engine or recommends consumer products can, with minimal adaptation, be repurposed to craft extremist manifestos or orchestrate disinformation campaigns at scale (Rickli & Liang, 2024, pp. 118–126).

Emerging technologies as force multipliers

AI and machine learning

Violent non-state actors are increasingly leveraging artificial intelligence to enhance propaganda precision and streamline recruitment. Machine-learning algorithms analyse user behaviour to identify ideological vulnerabilities, delivering tailored extremist content across linguistic and cultural boundaries. AI-powered chatbots simulate human interaction, sustaining engagement, radicalising targets, and retaining recruits at scale, at minimal cost and without operational fatigue (Kaur, 2025, pp. 28–29).

Drones and autonomous systems

Unmanned aerial systems provide low-cost, deniable capabilities for intelligence collection and precision strikes. Non-state actors have repeatedly deployed reconnaissance drones in conflict settings (Cohen & Freilich, 2024, pp. 143–148), while ISIS's drone program during the Mosul siege demonstrated how improvised devices can be converted into effective battlefield assets (Rickli & Liang, 2024, pp. 118–126) – capabilities traditionally associated with state militaries.

Encryption and the Dark Web

Encrypted communication platforms such as Signal, WhatsApp, and Telegram shield operational planning from interception, while dark web services facilitate the procurement of explosives, malware, and stolen data[6] (Kaur, 2025, pp. 28–29).

> **"Following post-2015 counterterrorism crackdowns, ISIS migrated significant elements of its propaganda and command infrastructure to anonymised networks, complicating intelligence and law-enforcement penetration.[7]"**

Case study: the Islamic State's cyber caliphate

The Islamic State (IS) provides one of the most instructive models for integrating emerging technologies into insurgent doctrine. Its "Dawn of Glad Tidings" Twitter application enabled decentralised, crowdsourced propaganda dissemination, generating tens of thousands of tweets daily before its removal (Sharma, 2020, pp. 368–391). Following significant territorial losses, IS migrated much of its infrastructure to encrypted communication channels, dark web hosting, and multilingual outreach, including content in Malayalam and Tamil, specifically designed to target Indian recruits.

[6] Cohen, M. S., & Freilich, C. D. (2024). Cyberterrorism. In Routledge eBooks (pp. 143–148).

[7] Sharma, A. (2020). Wilayat-e-Internet: Islamic State Cyber Caliphate. National Security, III–III, 368–391.

The 2014–2015 prominence of the @ShamiWitness account, operated by Indian engineer Mehdi Masroor Biswas, an IS propagandist, demonstrated how a single actor, leveraging anonymity and social media amplification, could become a global propaganda hub with millions of monthly impressions (Sharma, 2020, pp. 368–391). This convergence of technical proficiency, networked amplification, and ideological alignment exemplifies the asymmetric challenge of twenty-first-century counterterrorism.

India's vulnerability and strategic gaps

India confronts a dual strategic challenge: the rapid expansion of its digital economy and critical infrastructure has significantly widened the national cyber-attack surface, while the integration of doctrine and operation across cyber, intelligence, and kinetic domains remains incomplete.[8] Initiatives such as the Defence Cyber Agency (DCyA) and CERT-In represent important institutional progress, yet they lack systematic incorporation of AI-threat analysis, multilingual online monitoring capabilities, and scenario-based cyber wargaming into national security planning.

[8] UNIDIR. (n.d.). Building a More Secure world. United Nations Institute for Disarmament Research.

> **"Current global counterterrorism frameworks insufficiently account for adversarial AI risks, including data poisoning, model backdooring, and the exploitation of algorithmic bias in automated decision-support systems, vulnerabilities that could undermine both tactical operations and strategic stability.[9]"**

Conclusion: staying ahead of the curve

The weaponization of emerging technologies by violent non-state actors is no longer a speculative concern; it is an operational constant. From AI-optimised radicalization pipelines to drone-enabled urban warfare, the qualitative leap in capability has already transformed the threat landscape (Rickli & Liang, 2024). For India and its strategic partners, the imperative extends beyond merely hardening technical defences; it requires actively shaping the normative and legal frameworks that govern conduct in this contested digital battlespace.

The states that will prevail are those capable of integrating resilience, anticipatory threat analysis, and proactive diplomacy, denying adversaries not only access to critical systems but also the ability to control the narrative. This entails coupling robust cyber and AI governance with confidence-building measures, ensuring that technological innovation strengthens, rather than destabilizes, the international security order (Lubis, Muttaqin, & Nurwahidin, 2025).[10]

About the Author

Vibhuti Thapliyal is a Master's student in the History and Philosophy of Knowledge at ETH Zürich, where her research draws on science and technology studies (STS), philosophy, and the social sciences to examine emerging technologies, governance, and ethical decision-making. She previously worked with the Vivekananda International Foundation in the Technological and Scientific Studies division on projects focusing on AI-enabled warfare, autonomous weapon ethics, and the geopolitics of defence innovation. Her research interests include the "AI commander problem," accountability in autonomous military systems, and the governance of emerging technologies in hybrid and asymmetric conflict.

9 ENISA. (2023). AI Cybersecurity Guidelines. European Union Agency for Cybersecurity.

10 Lubis, H. A., Muttaqin, M. I., & Nurwahidin, N. (2025). The Cyber Proxy War: Non-State Actors Role in Global Geopolitical Competition. Journal Research of Social Science Economics and Management, 4(6), 815–828.


SYSTEM HACKED

“

The weaponisation of drones shows how cyberspace functions as both facilitator and force multiplier: it can connect intent with capability, and radicalisation with precise technical guidance

How open knowledge in cyberspace fuels drone weaponization

by **Lara Maria Guedes Gonçalves Costa**

Weaponizing drones in the digital age

In a widely reported case, a teenager in the United States (US) posted a YouTube video of a home-made drone firing a handgun and later modifying it with a functional flamethrower to roast a turkey. The video allegedly received backing from "HobbyKing," an online drone parts retailer, while the flamethrower's fuel pump came from Amazon.[1]

After allegations that he threatened to shoot people, he was expelled from a university. This incident illustrates the reality of improvised weapons production, driven by open knowledge and the diffusion of the dual-use technology know-how in cyberspace.

Improvised weapons are not new. The use of improvised explosive devices (IEDs) and small arms and light weapons (SALW) in both conflict and non-conflict settings is well documented.[2] However, uncrewed aerial vehicles (UAVs, or drones) have recently drawn renewed attention. In the ongoing war in Ukraine, civilians have adapted hobby drones into low-cost loitering munitions and reconnaissance tools.[3] In Central America, modified drones have been used for smuggling, surveillance, and attacks.[4] These examples reflect a surge in do-it-yourself (DIY) drone weaponization, a phenomenon amplified by the digital domain.

1 Ben Popper, "The Teenager Behind the Drone Gun Now Has a Drone-Mounted Flamethrower", The Verge, 8 December 2015.

2 See, for example: Matilde Vecchioni, "Unregulated Production: Examining Craft-Produced Weapons from a Global Perspective", UNIDIR, 20 June 2024.

3 Isabel Coles and Ievgeniia Sivorka, "Four Ways Ukraine's Drone Innovations Are Changing Warfare", The Wall Street Journal, 12 October 2024.

4 Henry Ziemer, "Illicit Innovation: Latin America Is Not Prepared to Fight Criminal Drones,", Center for Strategic and International Studies, 11 June 2025.

Guidance is only a click away. Online tutorials and open-source designs have been shaping perceptions and influencing youth in cyberspace, who are both prominent users of digital spaces and key figures in drone maker communities. As this trend grows, understanding these dynamics is critical to addressing the normalization of violence and the misuse of emerging technologies.

Open-source knowledge and the DIY community

UAVs are no longer prohibitively expensive and limited to a few states.

> **"The global market for commercial UAVs is forecast to grow from US$7.2 billion in 2022 to US$19.8 billion by 2031.[5]"**

In parallel, technological simplification has lowered the expertise required to operate or modify drones.[6] This accessibility has attracted not only armed groups and criminal networks, but also lone actors eager to explore and execute plots with information found online.

Instructions on building or modifying drones are readily available across digital communities. "DIY Drones," the world's largest group of drone enthusiasts, counts over 80,000 members worldwide.[7] While many such networks promote responsible innovation, their open and unregulated nature makes oversight difficult. Contrary to the belief that weaponization content is confined to the Dark Web, tutorials on mainstream platforms like YouTube demonstrate how to attach payloads, including weapon systems, to consumer drones.

---

5 "Teal Group Predicts Worldwide Civil UAS Spending of $139 Billion Over the Next Decade in Its 2022/2023 UAV Market Profile and Forecast-Teal Group", Teal Group Corporation, 4 November 2022.

6 Gregory D. Koblentz, "Emerging Technologies and the Future of CBRN Terrorism", The Washington Quarterly 43, no. 2, 2 April 2020: pp. 185–86.

7 DIY Drones: The Leading Community for Personal UAVs, last accessed 2 August 2025.

Entertainment-oriented channels have also entered the space. Series of videos like "Game of Drones" showcase a team arming drones with flamethrowers, rockets, and paintball guns with live targeting systems.[8] Most creators claim these projects are non-commercial and experimental, yet they still represent improvised weaponization. Viewer comments, often from young people, praise these designs and express interest in replicating them. The unregulated sharing of such modifications online makes it easy to repurpose them for harm.

The cultural imaginary of drones and violence

To understand this phenomenon's impact on youth, it is crucial to look beyond the existence of DIY drone tutorials and examine why they attract so many viewers, particularly young people. The reasons youth might engage in crime or terrorist activity vary, but one often-overlooked driver is the growing culture of militarization performed and promoted through the digital domain.

Scholars have long noted how different media sources shape the way people see, think about, and justify violence.[9] While depictions of violence in entertainment are not a novelty (e.g., in movies), the weaponization of drones has introduced new ways of imagining and normalizing it. In popular video games such as "Battlefield" and "Call of Duty," drones are designed to match real-world specifications and operate as lethal tools in virtual combat.[10] App stores host games featuring lethal UAVs, including ones where players hunt people.[11] These games also feature narratives, such as depictions of how a US-China drone war would look in 2025.[12] Often presented without reflection or criticism, such entertainment frames UAVs as ordinary consumer products rather than tools of violence.

[8] Marque Cornblatt Productions, "Paintball Drone Gunship - a DIY Combat UAV from Game of Drones", YouTube, 10 June 2013
[9] See Vieira and Krcmar (2011), Dillio (2014), Dholakia and Reyes (2018).
[10] Roger Stahl, "What the Drone Saw: The Cultural Optics of the Unmanned War", Australian Journal of International Affairs
[11] Stahl, "What the Drone Saw," pp: 667-668.
[12] *Ibid.*, p. 666.

It is crucial to note that the interest of the DIY community in weaponizing drones is part of a "wider cultural imaginary,"[13] a shared set of ideas and expectations in society about what drones are for. Beyond fictional scenarios in video games, these imaginaries are also shaped by real-world state practices.[14] People are generally aware of the historical use of drones by the military (e.g., US lethal UAVs in counterterrorism), especially given freely accessible footage online. Police forces have also explored the idea of deploying drones for civil unrest, for example through the use of pepper spray-equipped UAVs.[15]

> **"Even when intended for non-lethal or controlled use, the operational deployment of UAVs by state authorities feeds into a cultural narrative that travels – people see it, absorb it, and replicate it."**

This influence is visible in the rise of "DIY warfare," rooted in the "maker culture" that values informal, peer-led learning.[16] In Ukraine, for example, soldiers have worked alongside young civilians to build drones and counter-UAV solutions.[17] This knowledge-sharing has crossed into the open, unregulated online space. Today, online videos show how to build "kamikaze drones," UAVs designed to self-destruct upon reaching a target.[18] One widely viewed online tutorial, produced by a former combatant with over one million YouTube subscribers,[19] has been widely praised by drone enthusiasts, including youth, turning state-sanctioned violence into part of everyday entertainment and hobbyist culture.

Moving youth away from harmful content

The weaponization of drones shows how cyberspace functions as both facilitator and force multiplier: it can connect intent with capability, and radicalization with precise technical guidance. This creates a growing transnational threat that bypasses traditional arms regulation. Addressing it requires a multi-stakeholder approach.

While online content moderation raises debates about free speech and privacy,[20] "freedom from fear" must also guide these discussions. This means recognising that regulating harmful content is as essential as protecting legitimate expression. The Global Internet Forum to Counter Terrorism (GIFCT), founded by major tech companies,[21] should continue and strengthen its work to reinforce platform accountability, including through the development of more robust moderation and reporting tools to address weaponization-related content, while safeguarding lawful innovation.

Governments should ensure that cyberspace governance discussions include the risks posed by open-source knowledge related to emerging technologies. This requires engaging a broad range of

[13] Anna Jackman, "Consumer Drone Evolutions: Trends, Spaces, Temporalities, Threats", Defense & Security Analysis 35, no. 4, 24 October 2019: p. 370.

[14] *Ibid.*, pp. 370-371.

[15] "Pepper-Spraying Drones Could Be Used on Unruly Crowds by Indian Police", The Guardian, 8 April 2015.

[16] Akshat Upadhyay, "Do-It-Yourself (DIY) Warfare: A New Warfighting Paradigm", Strategic Analysis 48, no. 1, 19 March 2024: p. 18.

[17] "On the Front Lines with Ukraine's Killer Drone Pilot", The Wall Street Journal, video published on YouTube, 14 October 2024.

[18] "How Are 'Kamikaze' Drones Being Used by Russia and Ukraine?", BBC News, 29 December 2023.

[19] Civ Div, "How I Make Kamikaze Drones", YouTube, 19 June, 2024.

[20] Maura Conway and Stuart Macdonald, "Introduction to Special Issue: The Practicalities and Complexities of (Regulating) Online Terrorist Content Moderation", Studies in Conflict & Terrorism, 19 June 2023, pp. 1–4.

[21] GIFCT is an NGO founded by Meta (formerly Facebook), Microsoft, YouTube and X (formerly Twitter) in 2017. See: GIFCT, "About," GIFCT Global Internet Forum to Counter Terrorism", last accessed 2 August 2025.

actors to develop preventive strategies and technical guidance. For example, following a mandate from the UN General Assembly and through a multi-stakeholder consultative process, the United Nations Institute for Disarmament Research (UNIDIR) created a self-assessment tool for states to evaluate their capacity to counter IEDs.[22] A similar model could address UAV threats, merging arms control principles with cyberspace governance to close existing policy gaps.

Partnerships among maker communities, universities, and the private sector should embed ethical guidelines and "safety by design" principles into drone development.

"Digital literacy programmes should equip youth with skills to critically evaluate online content, highlighting the legal and humanitarian consequences of weaponization."

Institutions can reinforce this by integrating ethics modules into the non-weaponization of knowledge, especially in engineering and science courses, fostering a generation of innovators aware of the risks and responsibilities of technological expertise.

[22] Bob Seddon and Alfredo Malaret Baldo, "Counter-IED Capability Maturity Model and Self-Assessment Tool", UNIDIR, 24 June 2020.

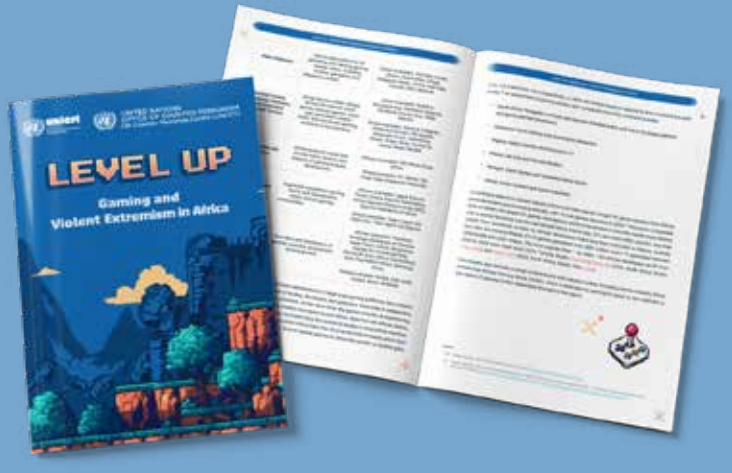
About the Author

Lara Maria Guedes Gonçalves Costa is a young professional working in the field of arms control and disarmament. She previously undertook the Graduate Professional Programme with the United Nations Institute for Disarmament Research (UNIDIR). Prior to that, she interned with the United Nations Office for Disarmament Affairs (UNODA) and the Stockholm International Peace Research Institute (SIPRI).

Lara also worked as an analyst at IBM BTO Business Consulting Services. She holds a master's degree in Security, Intelligence and Strategic Studies, jointly awarded by the University of Glasgow, Dublin City University and Charles University, and a bachelor's degree in International Relations and Area Studies from Jagiellonian University.

> **Governments should ensure that cyberspace governance discussions include the risks of open-source knowledge related to emerging technologies**

The new publication **Clicks, Links & Tricks, Oh My!** How Serious Organized Criminals Exploit Digital Trust Pathways examines how fundamental components of the Internet - domain names, uniform resource locators (URLs), and web traffic systems - are systematically manipulated. These elements, referred to throughout the report as "digital trust pathways", have become central enablers of a wide range of illicit activities. Strategically misused, they serve to facilitate, expand, and conceal criminal operations on a global scale.



Level Up – Gaming and Violent Extremism in Africa aims to deepen understanding of online harms in gaming spaces, particularly in the context of violent extremism. As gaming becomes increasingly social — especially through online mobile multiplayer titles with in-game chat — the potential for terrorist and violent extremist exploitation continues to grow.

The report **Right- and left-wing violent extremist abuse of digital technologies in South America, Africa and Asia**, jointly published by UNICRI and VOX-Pol, investigates the underexplored phenomenon of right- and left-wing violent extremist groups in the Global South and their abuse of digital technologies. As technology evolves at an unprecedented pace, violent extremist actors increasingly exploit digital platforms, posing complex and multifaceted threats to national and global security.

Download UNICRI publication

Traditional mutual legal assistance mechanisms often prove woefully inadequate in addressing such transnational complexity

Law beyond borders: the UN Cybercrime Convention navigating legal challenges in cross-border digital threats

by **Prof. Dr. Vladimir Aras**

In December 2024, through resolution 79/243, the United Nations General Assembly unanimously adopted the UN Convention against Cybercrime, marking a watershed moment in international criminal law.[1] This landmark treaty, the first comprehensive global cybercrime framework in over two decades, represents both humanity's collective recognition of digital threats and our evolving capacity to address them through multilateral cooperation.

The borderless challenge

Cybercrime has fundamentally transformed the criminal landscape, creating a paradox where the most sophisticated crimes occur in a realm that knows no borders, yet must be prosecuted within systems built on territorial sovereignty. Today's cybercriminals and threat actors operate with unprecedented impunity, exploiting jurisdictional gaps, regulatory inconsistencies, and the growing speed differential between technological innovation and legal adaptation.

Consider the anatomy of a modern ransomware attack: perpetrators in one jurisdiction infiltrate systems in another, encrypt data stored across multiple countries, demand payment through cryptocurrency networks spanning the globe, and launder proceeds through digital exchanges in yet other territories.[2] Traditional mutual legal assistance mechanisms often prove woefully inadequate in addressing such transnational complexity.

The challenge extends beyond mere logistics. Digital evidence is inherently volatile – easily altered, deleted, or relocated across servers worldwide within minutes.[3] The asymmetry between

[1] United Nations General Assembly, "United Nations Convention against Cybercrime; Strengthening International Cooperation for Combating Certain Crimes Committed by Means of Information and Communications Technology Systems and for the Sharing of Evidence in Electronic Form of Serious Crimes", December 2024.

[2] Chainalysis Team, "OFAC Sanctions Fraud Network Funding DPRK Weapons Programs", August 2025.

[3] Aurang Zaib Ashraf Shami, "Cybercrime and Digital Evidence: investigating the challenges and opportunities in prosecuting cybercrime and handling digital evidence", Research Consortium Archive, August 2025.

the speed of cybercrime and the pace of international legal cooperation creates a systematic advantage for criminals, who can outmanoeuvre law enforcement constrained by bureaucratic processes and sovereignty concerns.

A global response takes shape

The UN Cybercrime Convention emerges against this backdrop as an ambitious attempt to harmonize international responses to digital threats. Building upon the normative foundations laid by the 2001 Budapest Convention, the UN treaty attains an unparalleled degree of international legitimacy, evidenced by its unanimous endorsement by all 193 Member States.[4]

The Convention's architecture reflects hard-won compromises between competing visions of cybersecurity governance. It establishes standardized definitions for core offenses – illegal access, cyber fraud, online child exploitation – while creating mechanisms for rapid international cooperation through a 24/7 network system. States parties must criminalize illegal access to ICT systems; illegal interception of non-public transmissions of electronic data; interference with electronic data or ICT systems; misuse of devices that facilitate the commission of such offences; ICT system-related forgery; ICT system-related theft or fraud; offences related to online child sexual abuse or child sexual exploitation material; solicitation or grooming for the purpose of committing a sexual offence against a child; non-consensual dissemination of intimate images; and laundering of proceeds of the above offences. Most significantly, it emphasizes prevention strategies, technical assistance and capacity building, recognizing that cybersecurity is only as strong as its weakest link.

The treaty's provisions on international cooperation represent its most innovative elements. Traditional extradition processes, which can take years, are supplemented by streamlined evidence-sharing mechanisms designed to preserve digital evidence before it disappears. The Convention mandates that States designate competent authorities available 24/7 to respond to urgent requests for assistance, acknowledging that cybercrime operates on Internet time, not bureaucratic schedules. In addition, the Convention establishes a comprehensive multilateral framework designed to facilitate cooperation in transnational criminal matters, encompassing investigative activities, prosecutorial coordination, mechanisms for asset recovery, and the conduct of judicial proceedings.[5]

Opportunities and tensions

> **“The significance of the Convention resides not solely in the normative duties it codifies, but also in its capacity to cultivate a transnational community of practice in the prevention and prosecution of cybercrime.”**

By operationalizing States’ positive obligations under international law, the instrument reinforces the principle of due diligence in cyberspace, obliging States to adopt effective measures to prevent, investigate, and punish transnational offences facilitated by digital technologies.[6] In this sense,

[4] Hanoi Convention, “Official Statement”.

[5] Do Viet Cuong and Nguyen Quang Ha, “Hanoi Convention against Cybercrime: a milestone for the World and Vietnam”, Vietnam Law Magazine, April 2025.

[6] Carmen Eloísa Ruiz López, Vladimir Barros Aras, “A ação penal como um remédio efetivo para a defesa de direitos humanos: uma visão a partir da jurisprudência das cortes regionais”, Revista jurídica da presidência, March 2024.

the Convention not only harmonizes substantive and procedural norms but also institutionalizes mechanisms of mutual legal assistance and cross-border cooperation, particularly in the areas of evidence preservation, information-sharing, and judicial cooperation. The result is a normative framework that strengthens both horizontal collaboration among States and vertical accountability to international legal standards, thereby contributing to the consolidation of an integrated regime of global cyber governance. By establishing common standards and facilitating knowledge transfer from technologically advanced nations to developing countries, it promises to level the playing field against transnational criminal networks.

Yet the treaty's ambitions are tempered by significant concerns. Critics, including several civil society organizations, warn that vague language could criminalize legitimate security research and expand state surveillance powers without adequate safeguards.[7] The Convention's broad definitions–particularly those relating to "electronic data" and "information and communications technology system" – could create a framework susceptible to authoritarian abuse. I do not consider this to be accurate, given that the Convention is grounded in universally recognised human rights principles.

But some human rights issues are particularly troubling. While the treaty references international human rights law, it largely defers substantive protections to national legislation. This creates a dangerous precedent where authoritarian regimes could potentially use the Convention's cooperation mechanisms to pursue dissidents, whistleblowers, or journalists under the guise of cybercrime enforcement.

The path forward

The Convention's true test will come not in its formal ratification – expected to occur rapidly given its unanimous adoption – but in its implementation across diverse legal systems and political contexts. Following an extended negotiation process, the Convention was made available for signature on 25 October 2025 during a ceremony in Hanoi, Viet

[7] Kate Graham-Shaw, "New U.N. Cybercrime Treaty Could Threaten Human Rights", Scientific American, August 2024.

Nam, and will remain open for signature at United Nations Headquarters in New York until 31 December 2026. It is set to enter into force upon the deposit of the fortieth instrument of ratification or accession, after which its implementation will be subject to review by the Conference of the States Parties.

Success will require a careful balance between security imperatives and rights protections, international cooperation and sovereignty concerns, technological innovation and legal certainty. The Convention must evolve beyond a mere legal instrument to become a framework for ongoing dialogue and adaptation as digital threats continue to evolve. The stakes could not be higher. As digital technologies become increasingly central to economic activity, social interaction, and democratic governance, the failure to establish effective international cooperation against cybercrime threatens not only individual security but the integrity of the digital commons upon which modern society depends. We trust that the Convention will not only enter into force without delay, but will also secure wide ratification, thereby consolidating its authority as a universal instrument of international law.

Conclusion

The UN Cybercrime Convention represents both an achievement and a beginning. It demonstrates unprecedented global consensus on the need for coordinated responses to digital threats while highlighting the persistent tensions between security and freedom in the digital age. Its ultimate success will depend not on the elegance of its legal architecture but on the wisdom and restraint with which it is implemented by the international community, especially by prosecutors and courts across the globe.

As we stand at this inflection point, the Convention offers a framework for international cooperation that, if properly implemented with robust human rights safeguards, could significantly enhance global cybersecurity.

The choice before us is clear: we can use this moment to build a more secure and rights-respecting digital future, or we can allow legitimate security concerns to justify the erosion of the freedoms that make the digital revolution worthwhile.

> **“The UN Cybercrime Convention provides the framework; the global community must provide the wisdom to use it well.”**

About the Author

Vladimir Aras holds a PhD in Law, a Master's degree in Public Law, an MBA in Public Management, and has been a member of the Public Prosecutor's Office since 1993. He is a Senior Federal Prosecutor in Brasília, Adjunct Professor of Criminal Procedure and International Law (UnB), a member of the Attorney General's Office's (PGR) Cybercrime Task Force (GACCTI), former Head of the International Cooperation Unity within the PGR (2013-2017), and Founder of the Institute of Law and Innovation (ID-i).

unicri

University for Peace

MASTER OF LAWS (LL.M.)

in Global Criminal Justice and Accountability

16 November 2026 - 25 June 2027, Turin (Italy)

Apply by 6 September 2026

Scan for more

Explore our Master of Laws (LL.M.) program at www.unicri.org and take the next step in your legal education

unicri

University for Peace

MASTER OF LAWS LLM

IN CYBERCRIME, CYBERSECURITY AND INTERNATIONAL LAW

The hybrid threat of cyber-terrorist groups: critical gaps in the international legal framework

by **Matteo Pastorella**

Cyber conflict presents challenges beyond conventional security paradigms. A key blind spot is the use of non-state groups as state proxies.[1] The blurred distinction between state-sponsored and non-state-sponsored activities complicates attribution, weakening legal and strategic responses, and enabling aggression with impunity.

Cyber conflict is inherently asymmetric. By outsourcing operations to ideologically driven actors, states maintain plausible deniability,[2] complicating the application of international law on attribution and responsibility. Without a clear and direct state digital footprint, countermeasures become difficult, creating a permissive environment for malicious activities.

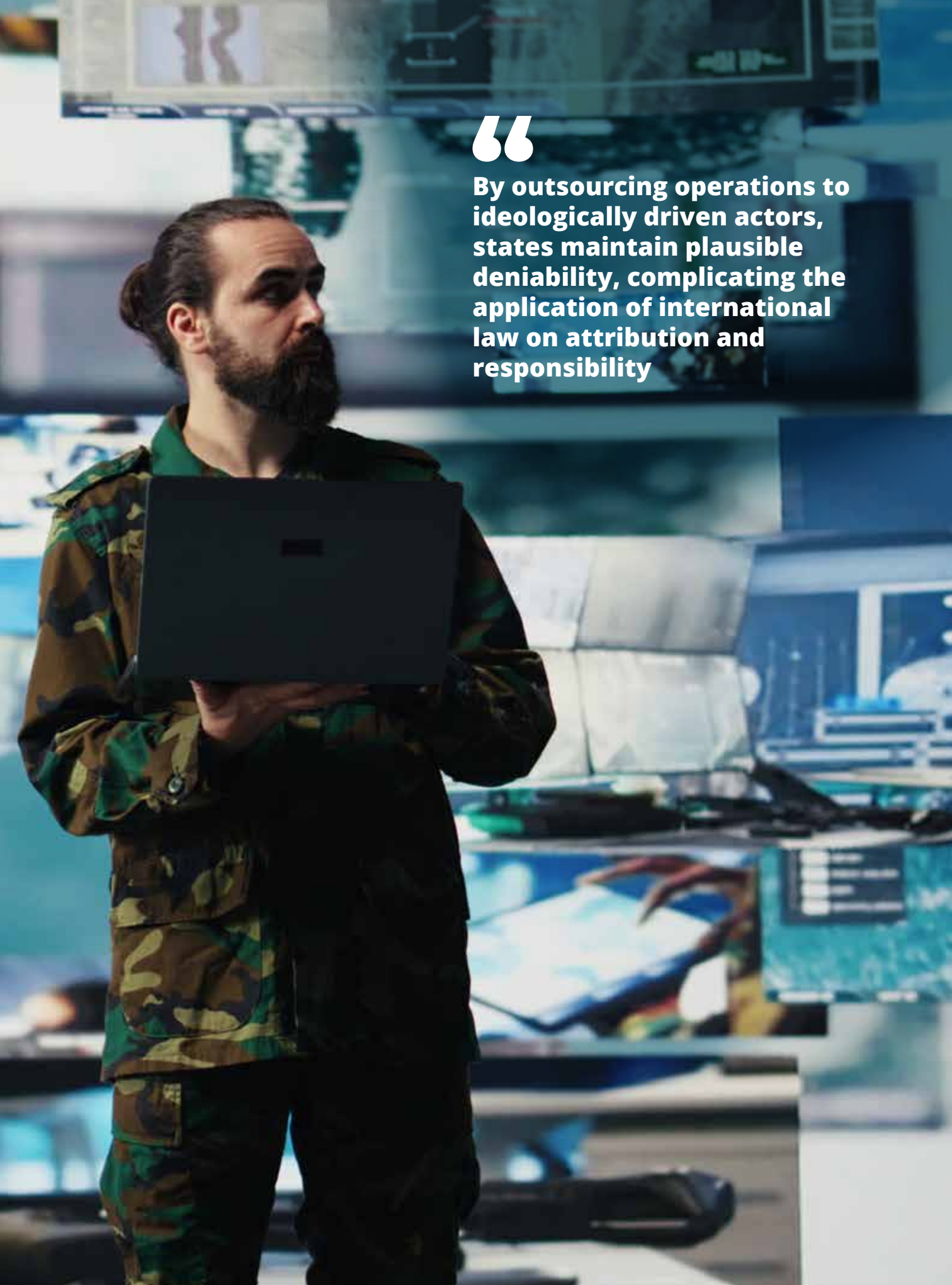
Hacktivist groups,[3] once decentralized and ideological, now access advanced capabilities through diverse levels of state sponsorship and support, carrying out malicious cyber activities as Advanced Persistent Threats (APTs) that rival military units. These APT campaigns are *advanced, deploying* logistical, technical and intelligence resources; *persistent, ensuring* long-term access; and a *threat, relying on* deliberate strategy with clear objectives. Unlike cyber-criminals seeking profit, cyber-terrorists pursue political aims, destabilizing societies and undermining security. This evolution requires re-categorizing cyber-terrorists within cyber threat intelligence: they are hacktivists with advanced capabilities whose campaigns must be treated as APTs.

This raises important challenges in the application of international law to cyber conflict. The reliance of states on APT actors is facilitated by a highly fragmented legal framework. Cyberspace is not in a legislative vacuum; it is governed by treaty and by customary norms applicable to the physical realm, including the United Nations Charter.

1 Katharina Kiener-Manu, Cybercrime Module 14: Key Issues – Cyberwarfare, *UNODC (n.d)*.

2 Justin Key Canfil, The illogic of plausible deniability: why proxy conflict in cyberspace may no longer pay, *Journal of Cybersecurity*, vol. 8, No. 1 (2022).

3 Check Point Software Technologies, What is Hacktivism? *Check Point Cyber Hub* (n.d.).



By outsourcing operations to ideologically driven actors, states maintain plausible deniability, complicating the application of international law on attribution and responsibility

"The central issue is not whether international law applies to cyberspace but how it should be applied."

Legally, the two concepts of attribution and the prohibition on the use of force (UN Charter, Article 2[4])[4] remain among the most contentious in defining a cyber-attack. Before attribution, it must be determined whether a cyber-attack is malicious – that is, a cyber operation breaching the prohibition on the use of force – since only that would justify proportional countermeasures.

As affirmed in the International Court of Justice's 1996 advisory opinion, Legality of the Threat or Use of Nuclear Weapons,[5] core legal principles apply regardless of the weapons employed – including cyber means and methods. Consequently, a cyber-attack may trigger the right of self-defence under Article 51[6] of the UN Charter and, theoretically, under Article 5 of NATO's North Atlantic Treaty. The 2007 Estonia cyber-attack[7] illustrated this possibility. Furthermore, the 2021 NATO Summit[8] confirmed that cyber-attacks could trigger a NATO response[9], thereby supporting an effect-based approach[10] that assesses consequences such as infrastructure destruction or loss of life to determine equivalence to traditional military attacks.

Once a cyber operation has been determined to be malicious, attribution is then analyzed across three independent dimensions: technical, legal, and political. Technical forensic analysis relies on indicators of compromise[11] and tactics, techniques, and procedures[12] of malicious actors. Yet, APT proxies and cyber techniques, such as Internet Protocol (IP) spoofing, hinder technical attribution. Consequently, legal attribution under international law is equally complex. Of the three dimensions, political attribution often takes precedence, limiting responses to media statements and diplomatic measures.

"While malicious acts – especially against critical infrastructure – are widely condemned, so-called naming and blaming often leads to uncertain results in today's polarized context."

Modern cyber conflict requires merging legal frameworks with new governance tools. Article 2[4] of the UN Charter, along with the principles of attribution, provides a legal basis but remains insufficient. For instance, although the Tallinn Manual[13] offers valuable guidance, it lacks binding enforcement mechanisms; similarly, the Schmitt Test,[14]

4 United Nations, Repertory of Practice of United Nations Organs: Supplement No. 7, *Volume I, Article 2(4)*.

5 International Court of Justice, Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 8 July 1996, ICJ Reports 1996.

6 United Nations, Charter of the United Nations, *Chapter VII, San Francisco, 1945*.

7 Eneken Tikk, Kadri Kaska and Liis Vihul, Analysis of the 2007 Cyber Attacks Against Estonia from the Information Warfare Perspective,Tallinn: Cooperative Cyber Defence Centre of Excellence, 2008.

8 North Atlantic Treaty Organization, Brussels Summit Communiqué, *NATO* (14 June 2021).

9 Tomas Minarik, *Cyber Attacks and Article 5: A Note on a Blurry but Consistent Position of NATO*, Tallinn: Cooperative Cyber Defence Centre of Excellence, 2010.

10 Henry Farrell and Charles L. Glaser, The Role of Effects, Saliencies and Norms in US Cyberwar Doctrine, *Journal of Cybersecurity*, vol. 3, No. 1 (March 2017), pp. 7–17.

11 Microsoft, What Are Indicators of Compromise (IoC)? *Microsoft Security* (n.d.).

12 National Institute of Standards and Technology, Tactics, Techniques and Procedures, *NIST Computer Security Resource Center* (n.d.).

13 NATO Cooperative Cyber Defence Centre of Excellence, Tallinn Manual on the International Law Applicable to Cyber Warfare, Tallinn: NATO CCDCOE, 2013.

14 James E. McGhee, Cyber Redux: the Schmitt Analysis, Tallinn Manual and US cyber policy, *Journal of Law and Cyber Warfare*, vol. 2, No. 1 (2013), pp. 64–103.

which assesses whether a cyber operation constitutes the use of force, exposes persistent ambiguities in attribution and response mechanisms.

Beyond legal instruments, cyber diplomacy plays a vital role in mitigating escalation risks and fostering international cooperation. The EU Cyber Diplomacy Toolbox[15] promotes multilateral cyber norms, deterrence, and responsible state behaviour, helping to mitigate geopolitical cyber destabilization. Meanwhile, confidence-building measures,[16] pioneered by the Organization for Security and Co-operation in Europe, enhance transparency and trust. In parallel, the UN Group of Government Experts outlined 11 voluntary, non-binding norms[17] for state behaviour in cyberspace, reinforcing existing standards against the misuse of information and communication technologies.

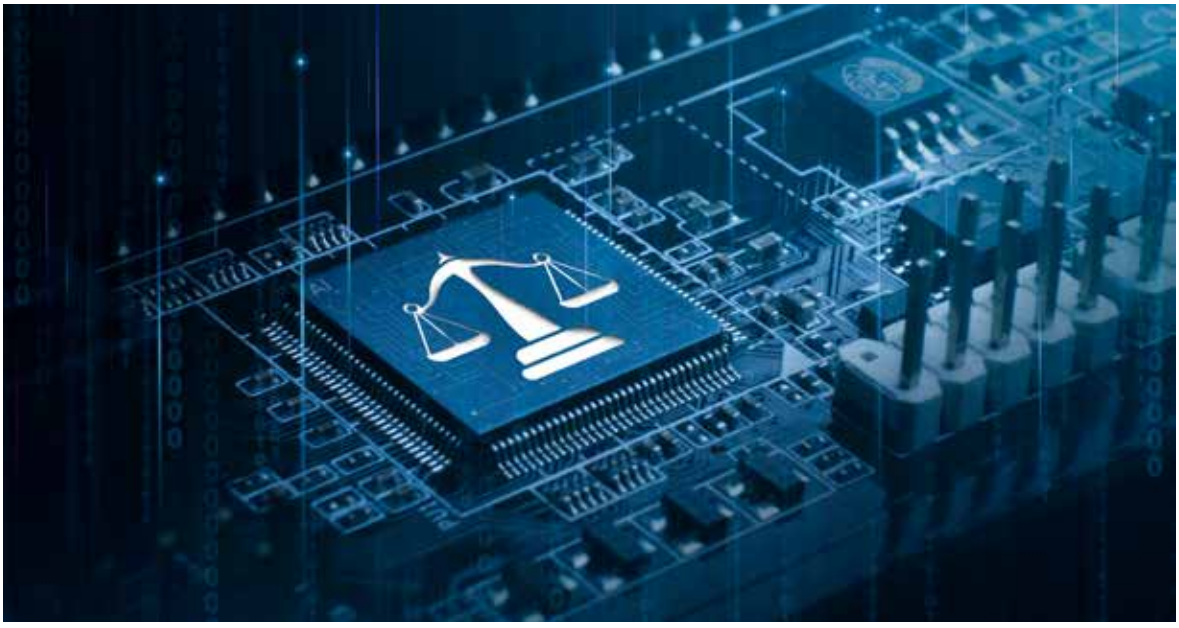
The systematic use of APTs remains a blind spot. Without effective attribution, state-sponsored aggression continues, undermining stability and sovereignty. The absence of a universal definition of cyber-terrorism – as seen by its omission in the 2001 Budapest Convention on Cybercrime[18] – complicates responses. Deterrence requires refining legal instruments, clarifying responsibility, and building enforcement. Without these measures, cyber aggression will persist, eroding security frameworks and destabilizing order.

[15] European Union, Cyber Diplomacy Toolbox, *European External Action Service* (n.d.).

[16] United Nations Office for Disarmament Affairs, Military Confidence-building Measures, *UNODA* (n.d.).

[17] United Nations General Assembly, *Resolution 70/237,* Developments in the Field of Information and Telecommunications in the Context of International Security, 23 December 2015.

[18] Council of Europe, *Convention on Cybercrime* (Budapest Convention), ETS No. 185, 23 November 2001.



About the Author

Matteo Pastorella is an advisor in international relations and cybersecurity. He holds several degrees, including degrees in Political Science, International Relations, and European Union Studies from LUISS Guido Carli and the University of Salzburg, as well as a Master's degree in Defence and Security from the University of Palermo. He has conducted research on defence and cybersecurity at specialized think tanks and parliamentary study centres, and has managed international projects at the Embassy of Colombia in Italy and the Italian Cultural Institute in Lima, Peru. He previously served within the Cybersecurity Unit of the Italian Ministry of Foreign Affairs and as a Cybersecurity Advisor in the framework of the G7 Task Force under the Italian Presidency of the Council of Ministers. He currently contributes to the Cyber Capacity Building Task Force at the Italian Ministry of Foreign Affairs. He has spoken at high-level institutional forums, including the Italian Parliament and the European Affairs Committee of the Chamber of Deputies, addressing issues related to cybersecurity, innovation, and European technological sovereignty. He is the author of several publications on cyber capacity building, hybrid threats, European strategic autonomy, and Latin American political processes, contributing to think tanks such as *Opinio Juris*, the Centro Studi Internazionali (CeSI), and the Istituto Affari Internazionali (IAI).

DECODING TRANSPARENCY

How to Foster Public Trust in Responsible AI Innovation in Law Enforcement

Public trust is the bedrock of legitimate, effective law enforcement. Its importance grows when law enforcement agencies adopt AI systems. Public attitudes towards AI in policing remain cautious, and trust in law enforcement agencies can strongly influence whether the public accepts new technologies. Therefore, it is up to law enforcement agencies to act effectively and, above all, fairly when making decisions about whether, when and how to adopt and implement AI systems.

DOWNLOAD THE PUBLICATION

The application of international humanitarian law to non-kinetic cyber operations

by **Mariam Salukvadze**

In recent decades, technological innovation has expanded the battlefield into the digital realm, where operations can incapacitate systems and undermine societal stability without a single shot being fired. Under International Humanitarian Law (IHL), applicability is triggered once the threshold of armed conflict is reached, as clarified by international jurisprudence such as the International Criminal Tribunal for the former Yugoslavia's (ICTY) Tadić Decision, in particular the Appeals Chamber's decision in Prosecutor v. Duško Tadić (Decision on the Defence Motion for Interlocutory Appeal on Jurisdiction, 1995), which states that "an armed conflict exists whenever there is a resort to armed force between States or protracted armed violence between governmental authorities and organized armed groups or between such groups" (para. 70), thereby clarifying the threshold for the existence of an armed conflict and providing key criteria - namely the intensity of violence and the organization of the parties - for distinguishing between international and non-international armed conflicts.[1] Building on this, the *Tallinn Manual* - a non-binding but influential reference expert study on the application of international law to cyber warfare, developed under the auspices of the NATO Cooperative Cyber Defence Centre of Excellence - suggests that cyber operations causing physical damage or injury could qualify as armed attacks.[2] While it is generally accepted that cyber operations conducted in the context of an existing armed conflict fall under IHL, the crucial issue is whether cyber operations, in the absence of kinetic force, can independently reach the threshold required to trigger the application of IHL.

Article 49(1) of Additional Protocol I defines an *"attack"* as *"acts of violence against the adversary, whether in offence or defence."*[3] This has traditionally been interpreted by the International Committee of the Red Cross (ICRC) and tribunals such as the ICTY

[1] Prosecutor v Dusko Tadic (Decision on the Defence Motion for Interlocutory Appeal on Jurisdiction) ICTY-94-1-AR72 (2 October 1995) para 70.

[2] Michael N Schmitt (ed), Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (2nd edn, CUP 2017) 415.

[3] Protocol Additional to the Geneva Conventions of 12 August 1949 and relating to the Protection of Victims of International Armed Conflicts (Additional Protocol I) (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3, art 49(1).

Technological innovation has expanded the battlefield into the digital realm, where operations can incapacitate systems and undermine societal stability without a single shot being fired

as requiring physical force.[4] However, the *Tallinn Manual* adopts a broader, effects-based approach, suggesting that the notion of *"acts of violence"* is not confined to physical harm but also includes actions with violent consequences.[5] Nevertheless, the experts involved in the Manual remain divided: some assert that only cyber operations resulting in physical damage constitute attacks under IHL, while others argue for recognizing intangible harms - such as severe data corruption - as sufficient.[6] The dominant kinetic equivalence theory, which links the definition of an attack to physical effects, has been criticised for its inability to capture the strategic objectives of contemporary cyber operations, which often aim to disrupt functionality rather than cause physical destruction of infrastructure.[7]

> **"The definition of an armed attack under IHL should encompass not only immediate physical effects but also broader, long-term consequences that may emerge over time, including economic disruption and environmental harm."[8]**

In the cyber domain, the impact of an operation can range from temporary inconveniences, such as a Distributed Denial of Service (DDoS) attack, to severe outcomes - such as triggering the physical destruction of critical infrastructure through manipulated digital commands.[9] Increasingly, modern conflict targets the functionality of infrastructure rather than its physical form, reinforcing the need for a broader interpretation of *"attack"* that includes the neutralization of key systems. However, expanding the definition of an *"attack"* to include non-physical effects also presents legal and ethical challenges. An overly broad interpretation risks undermining core IHL principles - namely, distinction and proportionality.[10] These principles are fundamental to maintaining the balance between military necessity and humanitarian protection. If every disruptive cyber operation is considered an *"attack,"* even those causing minor inconvenience, there is a danger that legal protections are applied too broadly or inconsistently, thereby diluting the normative clarity of IHL.[11] Nonetheless, this does not mean that cyber operations causing minor disruptions should be dismissed outright. While not all inconveniences meet the threshold of an attack, their cumulative effects - particularly when directed at critical civilian infrastructure - can escalate into significant harm. A

4 International Committee of the Red Cross, Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949 (Yves Sandoz et al eds, 1987) para 1880; Prosecutor v Pavle Strugar (Judgment) IT-01-42 (ICTY Appeals Chamber, 17 July 2008) para 270. See also Michael Bothe, Karl Josef Partsch, and Waldemar A Solf, New Rules for Victims of Armed Conflicts: Commentary on the Two 1977 Protocols Additional to the Geneva Conventions of 1949 (2nd edn, Martinus Nijhoff Publishers 2013) 329.

5 Michael N Schmitt (ed), Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (2nd edn, CUP 2017) 415.

6 Heather Harrison Dinniss, *Cyber Warfare and the Laws of War* (CUP, 2012) 199. Yoram Dinstein, T*he Conduct of Hostilities under the Law of International Armed Conflict* (3rd edn, CUP, 2016). See also Michael N Schmitt (ed), Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (2nd edn, CUP, 2017) 415-418. Michael Schmitt, 'Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework' (1999) 37 Colum Journal of Transnational Law 914-15.

7 Elizabeth Mavropoulou, 'Targeting in the Cyber Domain: Legal Challenges Arising from the Application of the Principle of Distinction to Cyber Attacks' (2015) 4(2) Journal of Law & Cyber Warfare 33.

8 Ido Kilovaty, 'Cyber Conflict and the Thresholds of War' (22 June 2021) forthcoming in Is the International Legal Order Unravelling? (David Sloss, ed, OUP 2022) 26.

9 Justina Nkechinyere Madubuike-Ekwe, 'Cyberattack and the Use of Force in International Law' (2021) 12 Beijing Law Review 636-637.

10 Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Additional Protocol I) (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3, art 48 and 51(5)(b). Kai Ambos, 'International Criminal Responsibility in Cyberspace' in Nicholas Tsagourias and Russell Buchan (eds), *Research Handbook on International Law and Cyberspace* (Edward Elgar Publishing 2015) 134.

11 Michael N Schmitt, 'Rewired Warfare: Rethinking the Law of Cyber Attack' (2014) 96(893) Int. Rev. of the Red Cross 205.

nuanced and context-specific approach is necessary, where the functional impact, strategic intent, and civilian consequences of a cyber operation are assessed holistically.[12]

In conclusion, cyber operations challenge the traditional frameworks of IHL, particularly with regard to the definition of *"attack* "and the protection of civilian objects. While IHL has historically been grounded in kinetic violence, the digital nature of cyber warfare calls for a broader, effects-based interpretation that accounts for both immediate and long-term disruptions to functionality. Nevertheless, this expansion must be carefully managed to avoid undermining foundational principles such as distinction and proportionality. A context-specific approach that considers strategic intent, the severity of harm, and civilian impact provides a more appropriate framework for evaluating cyber operations under IHL.

"As cyber warfare continues to evolve, legal norms must keep pace – grounded in clear state practice, strengthened by legal consensus, and guided by a commitment to safeguarding humanitarian values in the digital domain."

[12] Justina Nkechinyere Madubuike-Ekwe, 'Cyberattack and the Use of Force in International Law' (2021) 12 Beijing Law Review 646. Michael Gervais, 'Cyber-Attacks and the Laws of War' (2012) 30 Berkeley Journal of International Law 525, 525-531. See also Michael N Schmitt, 'Rewired Warfare: Rethinking the Law of Cyber Attack' (2014) 96(893) International Review of the Red Cross 204.

About the Author

Mariam Salukvadze is a legal professional specializing in criminal law, cybercrime, and digital human rights. In her previous role in law enforcement, she supervised investigators, and oversaw high-priority cases including cybercrime, ensuring the quality of investigations, and safeguarding human rights. She has contributed to the development of national policies on cybercrime and rights-based policing. As a Chevening Scholar, she holds a master's degree in criminal justice from Queen Mary University of London, awarded with distinction. She also contributes her expertise as a NonResident Fellow with the EU Cyber Diplomacy Initiative, an EUfunded programme that supports the Union's international cyber diplomacy efforts by advancing research, capacity building, and global cooperation to strengthen the rulesbased order and resilience in cyberspace.

Download Issue 5

unicri
United Nations Interregional Crime and Justice Research Institute

1540 Compass

ISSUE FIVE: INTERVIEWS WITH

PERMANENT REPRESENTATIVE OF PANAMA TO THE UN

Eloy Alfaro de Alba

DIRECTOR, SECRETARIAT OF THE ASIA-PACIFIC GROUP ON MONEY LAUNDERING

Mitali Tyagi

FORMER COORDINATOR OF THE 1540 GROUP OF EXPERTS

Jonathan Brewer

Justice by design: reimagining rights-based responses to cybercrime in Africa

by **Tina Power**

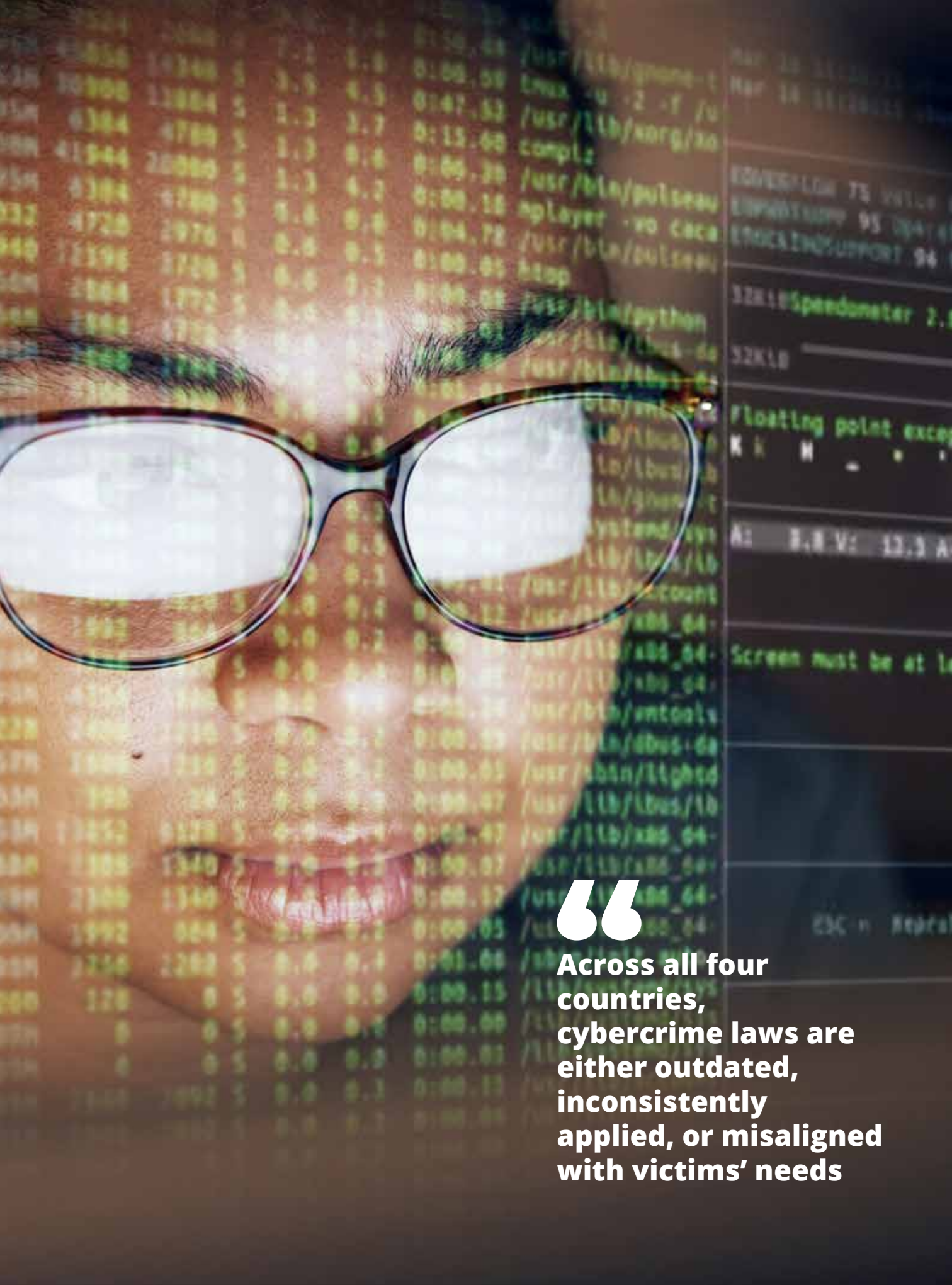
Earlier this year, UNICRI published a report on Access to Justice in the Digital Age: Empowering Victims of Cybercrime in Africa. Through this research UNICRI sought not only to identify the types of cybercrimes occurring in Africa, but more importantly to understand how people are being affected, and why so many remain without access to justice. Grounded in field visits, stakeholder consultations, and contextual research across East, Southern, and West Africa, the report unpacked the realties on the ground: first, a wide range of cybercrimes are prevalent across the regions; second, not all cybercrimes and online harms are treated equally, with responses often shaped by gendered dynamics; and third, deep structural barriers continue to obstruct access to justice, particularly for those most affected. This research underscored the urgent need to reimagine cybercrime not just as a technical threat, but as a justice issue, one that demands rights-based responses and systemic redesign.

Barrier to accessing justice

Having visited Uganda, Namibia, South Africa, and Sierra Leone, and despite differences in legal frameworks and institutional capacity, a common thread ran through each context: barriers to justice are widespread. Four key challenges emerged across the countries studied:

Outdated or inadequate legal frameworks

Across all four countries, cybercrime laws are either outdated, inconsistently applied, or misaligned with victims' needs. Namibia's Cybercrime Bill has been under discussion for nearly a decade, leaving a legal vacuum. South Africa has more advanced legislation, but frontline responders often lack clarity on how to record or respond to cybercrime reports. Sierra Leone's Cybersecurity and Crime Act is a welcome development, yet its effectiveness remains largely untested, with no finalised cases at the time of writing. In Uganda, stakeholders raised concerns that the law disproportionately targets government critics while offer-



“Across all four countries, cybercrime laws are either outdated, inconsistently applied, or misaligned with victims’ needs

ing limited protection or recourse for victims. These gaps not only undermine accountability but risk enabling misuse of legal tools, leaving victims without meaningful justice.

Capacity constraints

Law enforcement and justice systems across the region face significant resource and training deficits. Officers often lack the tools and technical knowledge to investigate and prosecute cybercrimes effectively. In Namibia, police training curricula exclude cybercrime and basic digital literacy, leaving officers ill-equipped to handle complex or sensitive cases, including online child sexual exploitation. Uganda faces similar challenges, with some police stations lacking even basic equipment like computers. The broader justice system also fails to integrate cybersecurity awareness. In South Africa, concerns were raised that judges are not receiving adequate training to properly navigate the nuances of cybercrimes. Sierra Leone has made strides through collaborative training efforts, but prosecutors still struggle to secure convictions due to weak investigations and insufficient evidence. These capacity gaps erode trust and deter victims from seeking help.

Knowledge gaps

Victims and frontline responders alike struggle to navigate existing mechanisms for redress.

> **“Many officers lack the specific skills and resources required to handle cybercrime cases, especially at the early stages of reporting where victims often seek immediate support.”**

For example, in South Africa, many police officers are not equipped with the knowledge to properly categorise and code cybercrimes that are reported. In Namibia underreporting is driven by a lack of knowledge; digital illiteracy; shame; perception of a lack of responsiveness from police and fears of re-victimisation. Although the Sierra Leone’s Cyber Security and Crime Act has been in place for several years, the police estimate that 70% of people remain unaware of its provisions or where to report cybercrime.

Low public awareness

Cybercrime is still poorly understood by the general public, particularly when it comes to personal or gendered harms such as online harassment and image-based abuse. Many people do not recognise these experiences as crimes, nor are they aware of their rights or available remedies. This lack of awareness contributes to underreporting and perpetuates impunity.

Designing for justice: six practical solutions

These challenges are widespread and contribute to persistent rights violations, but they are not beyond repair. With thoughtful redesign and systems that centre victims' needs, meaningful and accessible pathways to justice can be built. While some reforms demand deeper structural change, such as legislative reform, many practical improvements are within more immediate reach. Both must be pursued with urgency. UNICRI's report outlines a series of actionable solutions that can strengthen cybercrime response and advance justice for those most affected. Each of these reforms reflects a justice-by-design approach: practical, inclusive, and centred on victims lived realities.

Embark on law reform efforts: Cybercrime laws must be rights-based, clear, enforceable, and victim-centred. Vague definitions create loopholes and risk misuse, including silencing dissent. Strong legal frameworks help unlock budgets, guide enforcement, and build trust. Legislative clarity also enables better tools, training, and systems to protect victims.

Simplify reporting pathways: Victims need easy, trusted ways to report cybercrime. That means clear processes, trained responders, and user-friendly tools, like WhatsApp bots, hotlines, and online portals with step-by-step guidance. Reporting must be intuitive and accessible, especially for gendered harms.

Standardise cybercrime coding: Without consistent coding, cybercrimes can go untracked and unaddressed.

> **"Countries must develop clear, locally relevant systems to classify and record digital harms."**

These should be co-designed by justice sector actors, and backed by training. Public-facing versions should use plain language so victims can identify and report crimes.

Create practical SOPs: Police need clear, step-by-step guidelines for handling cybercrime. Standard Operating Procedures (SOPs) should cover everything from intake to investigation, with a focus on consistency, efficiency, and victim care. SOPs must be trauma-informed, rights-based, and easy to use.

Train justice sector stakeholders: Training is essential. Police, prosecutors, and judges must understand cybercrime, digital evidence, and victim-sensitive approaches. Programmes should be rights-based, focused on access to justice, practical, and tailored to emerging threats.

Launch a "Know Your Rights" campaign: Public awareness is power. A "Know Your Rights" campaign can help people recognise cybercrime, understand their options, and know where to seek help. It can tackle stigma, especially around gendered harms and promote digital safety.

These practical and realisable solutions are well within reach. It requires us to start off by simply shifting the narrative: cybercrimes must be understood not only as a technical or security issue, but as a human rights and justice issue.

It then requires some redesign. Across Uganda, Namibia, South Africa, and Sierra Leone, the barriers to justice are real, but they are not immovable. By embedding justice into the design of laws, systems, and institutions, we can shift from reactive enforcement to proactive protection. This means centring victims, simplifying access, and building capacity across the justice cycle.

> **"Justice by design is not a slogan, it is a blueprint for action. It calls on policymakers, practitioners, and communities to reimagine cybercrime responses as pathways to dignity, accountability, and inclusion."**

With urgency and collaboration, Africa can lead the way in crafting digital systems that protect rights, restore trust, and deliver justice where it's needed most.

About the Author

Tina Power is an Attorney of the High Court of South Africa and a Director of ALT Advisory - an African-rooted collective of public interest lawyers, researchers, and technologists working for positive social change. Tina works on advancing digital rights through research, policy reform, and strategic litigation across domestic, regional, and international contexts, having worked in over 25 countries. She regularly consults to various UN agencies and recently supported UNICRI with its report on Access to Justice in the Digital Age: Empowering Victims of Cybercrime in Africa. Tina has experience in human rights advocacy, research and training with a focus on access to justice, online harms reduction, and the promotion of equality, non-discrimination, and free expression both on and offline. Tina has worked closely with victims and survivors of cybercrimes, as well as activists, journalists, and international institutions. She holds an LLM in Human Rights Advocacy and Litigation from the University of the Witwatersrand and an MSc in International Human Rights Law from Oxford University.

Access to Justice in the Digital Age: Empowering Victims of Cybercrime in Africa

This new report explores how cybercrime is impacting access to justice in four African countries (Namibia, Sierra Leone, South Africa and Uganda) and offers a broader perspective on challenges and responses across the continent. It highlights the pressing need for national and regional responses that are inclusive, coordinated and evidence-based.

Download the publication

Digital rights and legal protection in unstable countries: An Iraqi youth view

by **Reman Mohammed**

Introduction

In today's digital world, online threats easily cross borders. In conflict-affected areas, people face heightened risks from online surveillance, disinformation, and privacy violations. Iraq is one such country, where weak legal protections, limited digital literacy, and fragile cybersecurity infrastructure leave communities, especially displaced people and youth, exposed to virtual threats that mirror their real-world vulnerabilities.

The reality on the ground

As a legal expert working in Iraq, I have witnessed how online threats deepen the struggles of already marginalized populations. Internally displaced people (IDPs), returnees, and young activists face serious risks, ranging from online harassment and identity theft to defamation and data breaches. Yet, Iraq's legal system remains ill-equipped to either address cybercrimes or hold perpetrators accountable.

Gaps in law and awareness

Iraq lacks a comprehensive legal framework for cybersecurity and digital rights. Critical issues like cyberbullying, online gender-based violence, and unauthorized surveillance are largely unaddressed in current laws. This legal vacuum endangers youth and human rights defenders who use online platforms to document abuse, advocate for justice, or organize their communities.

Humanitarian work and digital threats

In my role managing protection programs, I lead legal awareness sessions that often focus on documentation and housing rights. Increasingly, though, participants raise concerns about social media threats, scams targeting displaced persons, and cyberintimidation. These issues highlight the urgent need to incorporate digital protection into humanitarian legal assistance.

Critical issues like cyberbullying, online gender-based violence, and unauthorized surveillance are largely unaddressed in current laws

Youth as agents of change

Young people in Iraq and other fragile contexts are not just victims of digital threats, they are frontline defenders of justice. Through youth-led initiatives supported by the International Trade Centre (ITC), I have seen young leaders use digital tools to raise awareness, mobilize support, and promote rights. But they need targeted training in cybersecurity and legal literacy to safeguard these digital spaces.

Digital trust as critical infrastructure

In post-conflict recovery, trust in digital systems is as vital as rebuilding roads and schools. Cyber protections must be rooted in human rights, transparency, and accountability. A sustainable response to digital threats requires humanitarian strategies to include cybersecurity, especially for vulnerable youth.

Key steps orward

- Legal reform: Iraq must enact legislation that defines and penalizes cybercrimes while protecting free expression and privacy rights.
- Capacity building: Legal actors, NGOs, and youth leaders must receive training in digital safety, privacy law, and secure communication tools.
- Youth participation: Digital policymaking should reflect the lived experiences of youth, particularly in conflict zones.
- International support: Global agencies should integrate cybersecurity into humanitarian programming and provide technical support for digital protection systems.

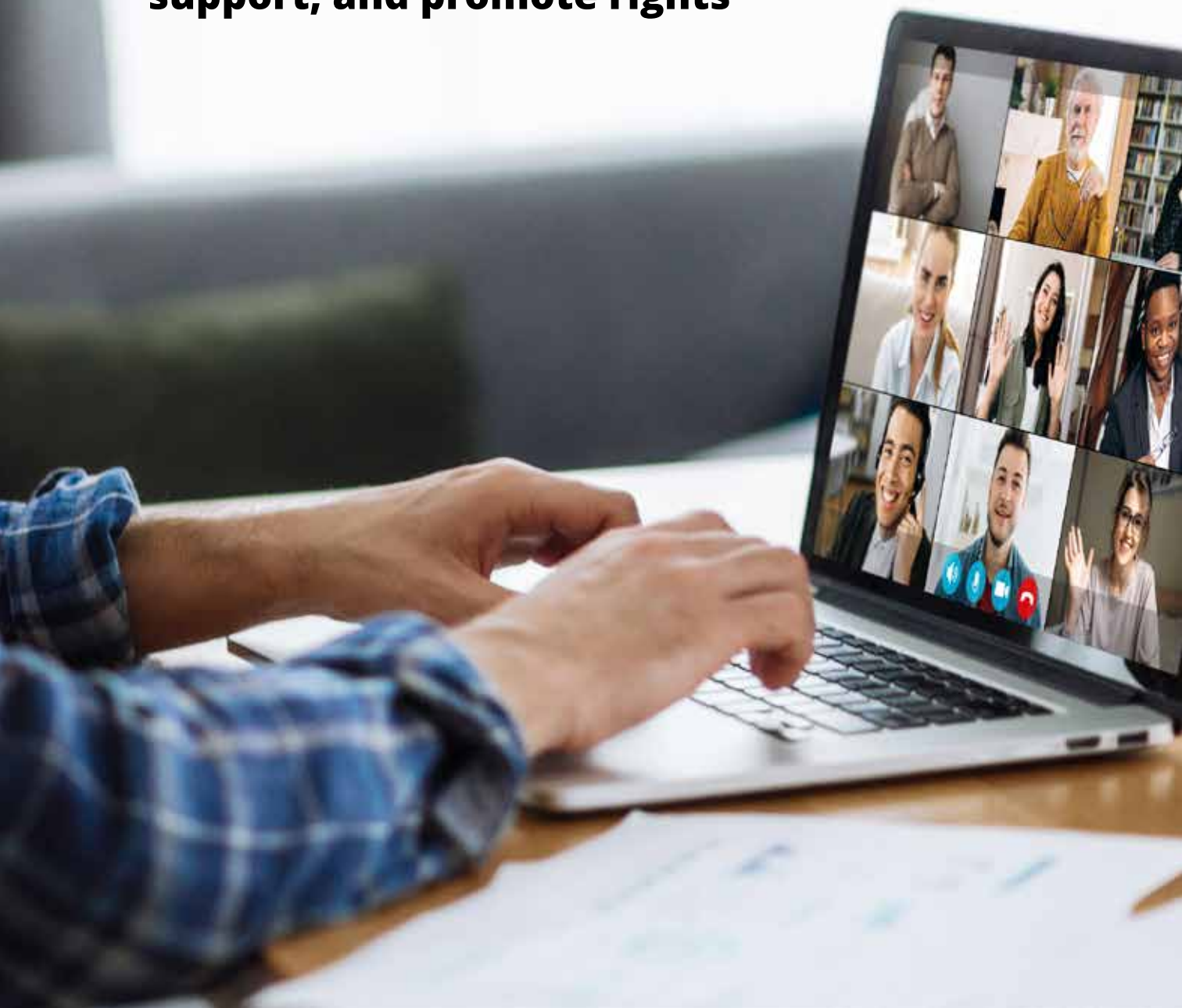
Conclusion

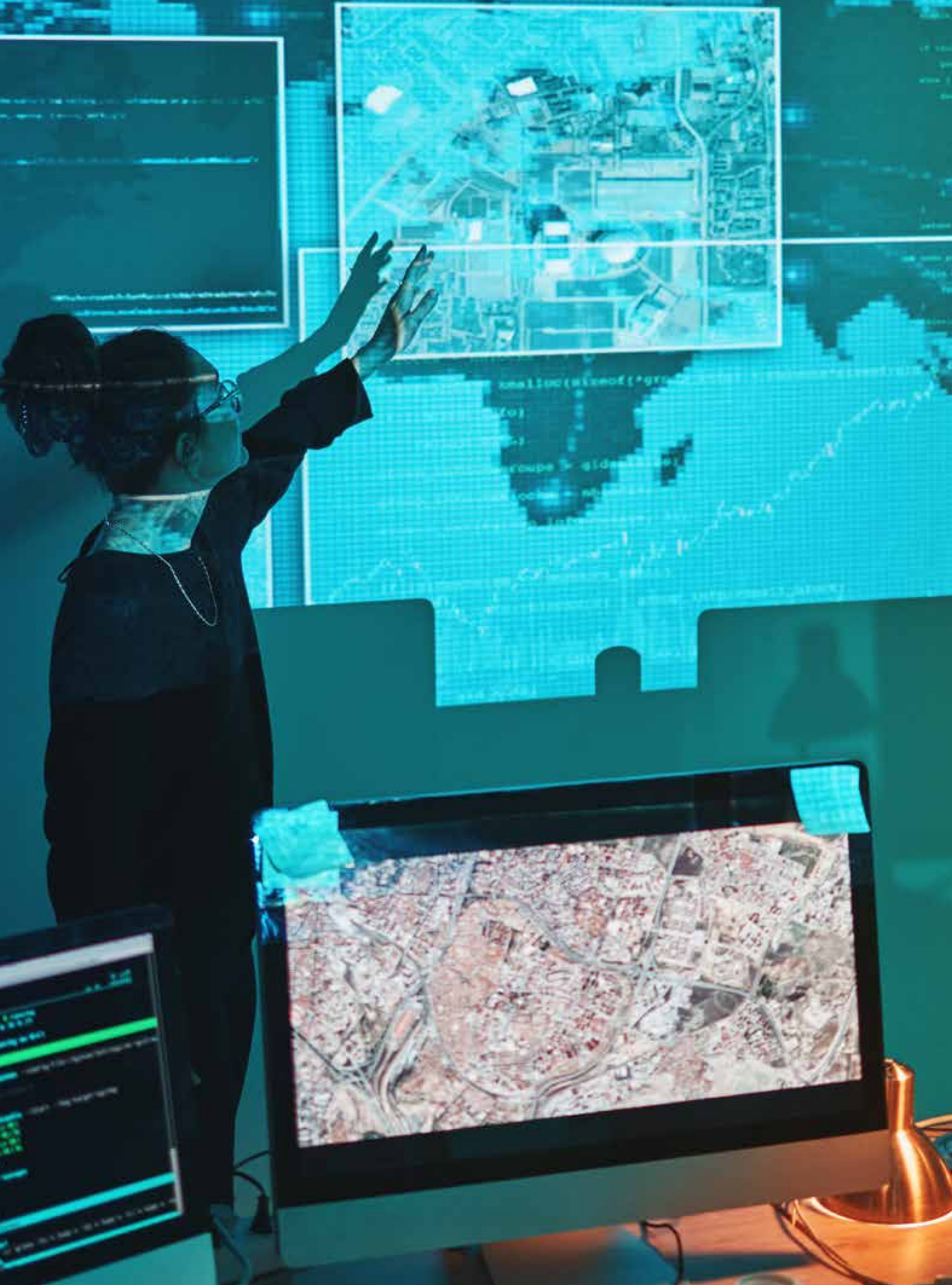
Online threats are not abstract or futuristic; they are immediate and linked to justice, human rights, and community resilience. Youth in Iraq and across the MENA region are ready to lead, but they need safe, legally protected digital spaces to do so. That is the new frontier of justice.

About the Author

Reman Mohammed is a senior Legal Expert and Humanitarian Program Manager with over 11 years of experience in protection, governance, and legal reform in conflict-affected areas. He currently serves as a Multi-Core Competency Program Manager with the Norwegian Refugee Council (NRC) in Iraq, where he leads multi-sectoral Protection and Information, Counselling and Legal Assistance (ICLA) interventions. Reman previously served as the National Legal Advisor for the Institute for International Law and Human Rights (IILHR), where he specialized in legislative drafting, justice sector reform, and policy development. With a B.A. in Law, a PMP certification, and an advanced background in International Humanitarian Law, he has a proven track record of managing large-scale donor portfolios and bridging the gap between human rights and technology to promote inclusive justice.

“Through youth-led initiatives supported by the International Trade Centre (ITC), I have seen young leaders use digital tools to raise awareness, mobilize support, and promote rights





Public–private partnerships as the scaffolding for safer digital spaces

by **Dr. Nagham El Karhili**

The screens we scroll and the streets we walk are part of the same ecosystem: harms that begin online can and do spill into the physical world and vice versa. As features, platforms, and communities proliferate, so do the vectors through which terrorist and violent extremist (TVE) messages travel — from recruitment and radicalization to coordination and the circulation of traumatic imagery. That reality makes clear that keeping online spaces safe is not solely an engineering problem for companies, nor only a law-enforcement task for States; it is a collective challenge that demands durable public–private partnerships, multistakeholderism, and cross-sector cooperation.

Public–private partnerships are one way this idea takes shape in practice. No single sector can prevent, mitigate, or respond to the misuse of digital spaces on its own. Industry can move quickly on technical fixes, but it needs government alignment, civil society insight, and academic expertise to ensure those fixes are responsible, rights-based, and contextually informed. The Global Internet Forum to Counter Terrorism (GIFCT) is one prominent example of this approach.[1] Originally an industry-led partnership, it evolved into a multistakeholder organization and was formalized in 2017 when major platforms agreed to share signals and research to reduce the online reach of violent extremists.[2] Those early cooperative steps — including the establishment of a cross-platform hash-sharing database (HSDB) — underscored that technical progress gains strength when supported by multistakeholder guidance.[3]

The March 15, 2019, Christchurch attacks made this reality tragically visible. What began as an offline act of mass violence was rapidly amplified online: the attacker livestreamed the massacre

[1] "Global Internet Forum to Counter Terrorism".

[2] Microsoft Corporate Blogs, "Facebook, Microsoft, Twitter and YouTube provide update on Global Internet Forum to Counter Terrorism", December 2017.

[3] GIFCT, "GIFCT's Hash-Sharing Database".

and posted a manifesto, both of which spread virally across platforms. Within 24 hours, companies reported removing millions of re-shares, a crisis that underscored two truths: viral harm can outrun single-company defenses, and coordinated, cross-platform response is essential. The attack spurred greater collaboration, pushing platforms not only to scale up incident response together, but also to bring governments and civil society more directly into the process.[4] Post-incident multistakeholder debriefs — including those convened by GIFCT — became a way to reflect, share lessons, and collectively strengthen preparedness for future incidents.

From these early evolutions, a practical lesson emerges: technical fixes are necessary but not sufficient. A reliable response architecture needs three mutually reinforcing elements: rapid, privacy-respecting technical signals; consultative governance to set rules of engagement; and external, rights-focused scrutiny to ensure fairness and legitimacy. GIFCT's trajectory illustrates this lesson well. Its HSDB, for example, provides perceptual "fingerprints" of known terrorist content so re-uploads can be detected and removed more efficiently. But what gets hashed and why is not, and should not be, a purely technical choice: inclusion criteria, taxonomy, and review processes are set through consultative working groups, regular taxonomy reviews,[5] and cross-sector dialogue so that operational tools reflect context and rights considerations as they scale.

Governance is equally important in public–private partnership models. GIFCT's Operating Board — composed of member technology companies — sits alongside an Independent Advisory Committee (IAC) made up of representatives from government, civil society, and intergovernmental organizations. The IAC provides external guidance on priorities, human rights concerns, regional dynamics, and transparency — a model that balances industry agility with independent scrutiny, while amplifying the voices of practitioners and experts who witness harms in different contexts around the world.

In practice, multistakeholder systems can deliver value in three key ways:

First: reducing information asymmetries through convening and knowledge exchange. Threat intelligence is often fragmented: platforms, governments, local civil society, and researchers each hold pieces of the same puzzle. Regular working groups, regional workshops, and threat-specific briefings knit these pieces together so that a public-health-style response can replace ad-hoc firefighting. Routine convenings also help small or regional platforms adopt best practices and prevent gaps that malicious actors may exploit. GIFCT's regional workshops are one example of how consistent convening builds shared situational awareness.

Second: translating operational signals into interoperable technical tools — responsibly. Hash-sharing is a clear example of how industry cooperation reduces re-uploads without requiring every company to build duplicate detection systems from scratch. But the utility of a shared technical tool depends on governance: who decides what is included, what criteria are used, how content is reviewed, and how to handle edge cases where context matters. Iterative taxonomies, inclusion criteria, and incident-response protocols — developed through consultative working groups and subject to independent review — are what make technical interoperability credible and defensible. GIFCT's taxonomy work and transparency reporting illustrate this balancing act.

[4] GIFCT, "The Incident Response Framework".

[5] GIFCT Working Group, "Broadening the GIFCT Hash-Sharing Database Taxonomy: An Assessment and Recommended Next Step", July 2021.

Third: anchoring responses in human rights and independent oversight. Rapid technical responses can reduce harm, but they can also chill legitimate expression or unevenly burden marginalized communities if safeguards are not embedded. Independent advisory mechanisms, human-rights due diligence, and external impact assessments are therefore central to effective public–private partnership. GIFCT's engagement with external experts, its IAC, and its human-rights assessment work demonstrate how accountability and operational speed can be reconciled.

Looking ahead, the tech and threat landscapes continue to shift. Generative AI, new content formats, and a diversified tech stack — from gaming and marketplaces to encrypted services and decentralized platforms — expand both the vectors of exploitation and the tools for mitigation. That duality means multistakeholder systems must accelerate knowledge exchange to ensure policy and engineering are informed by context, ramp capacity building for smaller platforms, and invest in interdisciplinary research that connects technical signals with sociopolitical dynamics. GIFCT's academic research arm, the Global Network on Extremism and Technology (GNET), is one example of how industry and academia can be bridged to advance this work.[6]

There are hard choices ahead: how to preserve open discourse while reducing the reach of violent narratives; how to equip smaller companies without encouraging automated over-removal; how to ensure incident response remains rapid without centralizing control. The only realistic way to nav-

[6] "Global Network on Extremism and Technology (GNET)".

igate those tradeoffs is together. Multistakeholderism — not as a slogan but as a practice of shared rules, independent scrutiny, regional engagement, and transparent feedback loops — is the best path we have to design online spaces that are resilient, rights-respecting, and responsive.

In short: terrorists and violent extremists will keep adapting; so must we. If public safety, human dignity, and democratic values matter, then industry, governments, civil society, and academia must keep designing and testing solutions together. The task is ongoing, but the lesson is timeless: the screens we scroll and the streets we walk are inseparable — and when they intersect, our responses should too.

About the Author

Dr. Nagham El Karhili is the Membership and Programs Senior Lead at the Global Internet Forum to Counter Terrorism (GIFCT). She oversees the full membership cycle, recruiting and mentoring new tech platforms and facilitating ongoing engagement with industry, government, and civil society partners. Her work ensures GIFCT's membership and programming strategies advance its mission to combat terrorism and violent extremism online while upholding human rights and addressing diverse multistakeholder needs. Before joining GIFCT, Dr. El Karhili served as Program and Research Manager at the Horizon Forum, a think-and-do tank focused on hate-funding in philanthropy, where she led investigative projects and convenings that informed policy recommendations and sector best practices. She holds a PhD in Communication from Georgia State University, where she was a Presidential Fellow at the Transcultural Conflict and Violence Initiative, examining violent extremism, organizational religious identity, and civil society resilience. She has taught media, religion, and peacebuilding at Georgia State University and Purdue University, and has also earned a BS and an MS from the University of Louisiana at Lafayette.

FREEDOM FROM FEAR
M A G A Z I N E

The New Criminal Code

Deciphering Emerging Threats in Cyberspace

DOWNLOAD ISSUE 20

"Cyberspace has become a defining arena for contemporary crime, conflict and security"

in FOCUS

The links between terrorism and organized crime in Mali

by **Adama Mamadou Ballo, Ibrahim Ahmadou Dicko, Ibrahim Traore**

> **"Over the past decade, numerous analyses have highlighted the growing fragility of the Sahel region, exacerbated by the expansion of terrorist and criminal activities."**

In West Africa, terrorist groups have intensified their operations and expanded their influence, threatening not only the Sahel but also coastal countries. At the same time, criminal networks play a major role in regional dynamics, exploiting economic and logistical opportunities for illicit purposes.

Based on desk research and interviews[1] with 39 key stakeholders across 6 regions of Mali, this study sheds light on complex dynamics in which the interdependence between crime and terrorism is both economic and logistical. These relationships fluctuate between cooperation and rivalry, influenced by strategic interests and local contexts. The observed criminal dynamics reveal a deep entanglement between organized crime and terrorism, fuelled by the isolation of certain regions, their cross-border positioning, weak state presence and socioeconomic precariousness.

Current trends indicate a geographic expansion of illicit activities into new areas, a diversification of trafficking (drugs, weapons or cultural property) and a growing convergence between terrorist groups and criminal networks, sometimes described as "coopetition" (cooperation and competition).

Beyond a strictly security-based approach, combating these phenomena requires a comprehensive effort to reduce population vulnerability and strengthen local governance. Rapid-impact actions (road improvements, youth reintegration programmes, vocational training, entrepreneurship support for women) must go hand in hand with strengthened support for local authorities, the justice system and security forces in order to build trust and reinforce the rule of law. The diversity of contexts between the north and the south of the country also calls for differentiated and tailored responses.

[1] Based on desk research and interviews conducted in 2024.

Key findings

The elements described below were identified through a combination of sources including desk research, meetings with specialized magistrates and in-depth interviews involving a total of 39 key actors. These included representatives of the State and local authorities, community and religious leaders, civil society actors, artisanal gold miners, self-defence militias, transport organizations or drivers' unions, and representatives of the chamber of commerce.

Interviews

Interviews were conducted in six administrative regions of Mali involving some of the areas most directly impacted by terrorism and organized crime, as well as less affected ones. They took place in nine major cities and included:

- Administrative authorities (representatives of the State and local authorities).
- Community and religious leaders.
- Civil society actors (e.g., artisanal gold miners, representatives of youth and women's organizations, transport organizations and the chamber of commerce).

Terrorism

Cattle theft is a widespread phenomenon that particularly affects herders and represents a lucrative activity for criminal and terrorist groups. Livestock buyers in major cities are sometimes involved in these networks.

The regions of Timbuktu and Gao are the main hubs for migrant smuggling, with thousands of migrants arriving each month and engaging with smugglers to reach border areas irregularly. The activities of terrorist armed groups and illicit migrant smuggling remain largely distinct; the link between the two is mainly in the form of extortion or a transaction resembling a right of passage through controlled territory.

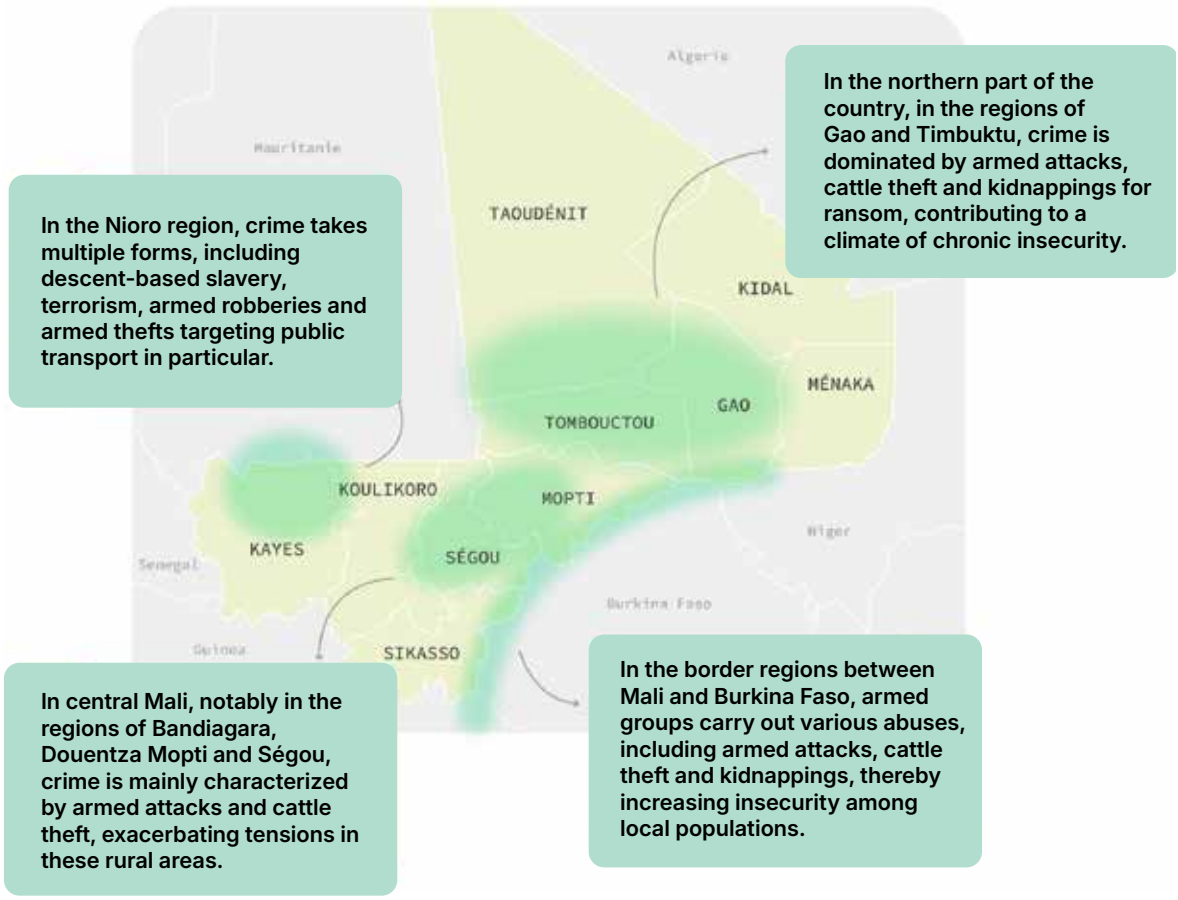
“The illicit arms trafficking market has expanded in response to increased demand from several actors: terrorist groups; militias; community self-defence groups such as Dan Na Ambassagou; and local communities seeking to protect themselves or to fuel intercommunal conflicts.”

Trafficking in cultural property is a phenomenon that has intensified in Mali since 2012 with the looting of archaeological sites, notably those in Djenné-Djeno and in the Central Delta. Thousands of objects have been removed and sold.

Kidnapping is a lucrative activity for terrorist groups who target civilians, humanitarian workers, journalists, military personnel and civil servants. Criminal networks often facilitate the capture, detention and transport of hostages, as well as the management of ransom payments. Several court cases have highlighted collaboration between these networks and militias or terrorist groups. This activity is growing rapidly and numerous investigations are ongoing.

“Terrorist and criminal groups often establish a form of forced zakat (a payment traditionally for charitable and religious purposes), requiring communities to pay a sum calculated according to the number of people or livestock”

(typically around CFA 5,000 per head). In the north, zakat is generally paid in animals or cash, while in central and southern areas the levy may take the form of agricultural produce. Armed groups sometimes entrust the storage and resale of animals to



intermediaries responsible for converting livestock into cash, in exchange for a payment of between CFA 200,000 and 500,000 per operation. Actors involved in customary governance may also participate in the collection of these levies, while religious leaders are generally not involved in this dynamic. "Taxes" may also be imposed on specific activities, such as trade, artisanal gold mining or attending Christian church services. At artisanal mining sites, a weekly tax (ranging from CFA 5,000 to 250,000) is imposed to access the site, while merchants, transporters and food vendors also pay fees to operate there. Terrorist groups have succeeded in taking control of several gold mining sites in the Gao and Kidal regions, where they can recruit new members and access explosive materials.

Taxes are paid to authorize access to the various sites in exchange for protection. In some of them, miners are required to pay to be protected, to gain access to water and food, and to obtain the right to establish camps at the gold extraction sites. Gold extracted in the Kidal and Gao regions is transported overland to several destinations, mainly Burkina Faso and Niger, and southward to Bamako, often to be exported onward to Southwest Asia.

Terrorist groups use shell companies, parallel markets and complex banking transactions to launder

money derived from criminal activities. The Financial Intelligence Unit (Cellule nationale de traitement des informations financières) has worked with other relevant Malian authorities to trace suspicious financial flows. Individuals involved in the financing of terrorism have been arrested in Bamako, Kayes and Ségou, and several of their assets have been seized as part of ongoing investigations.

> **"Mali constitutes a major transit point for cocaine trafficking between Europe, South America and other regions."**

Terrorist groups profit from these routes by imposing taxes on shipments or using drugs as a means of payment to acquire weapons. Currently, more than a dozen ongoing judicial cases are seeking to clarify possible links between drug trafficking and terrorism.

Analysis of the links between organized criminal groups and terrorist groups

BOX 1:

Illicit trafficking and criminal activities often serve as a point of convergence between the two types of groups, with a strong economic interdependence. Terrorist and criminal groups collaborate in specific activities such as cattle theft, crop destruction, kidnappings for ransom and vehicle hijackings in order to ensure an essential financial and logistical flow for their operations.

Box 2:

The nature of the relationship adapts to the context. The majority (60 per cent) of respondents in Gao observed the emergence of a form of "coopetition" between terrorists and criminals, combining rivalry and collaboration depending on shared interests (financing, logistics, protection). By contrast in the south, which is not as affected by both phenomena, competition and confrontation predominate making cooperation more discreet.

Vulnerability and resilience

Vulnerability factors

Geographic characteristics can exacerbate vulnerability to criminal activities. Strategic transit points for illicit trafficking, the vast extent of porous borders and the isolation of certain desert areas create favourable conditions for the development of these activities while making surveillance particularly complex. Local carriers involved in the transit of illicit goods, such as drugs or weapons, as well as local populations identifying secure routes, become essential links in the logistical chains of criminal networks. This collaboration creates interdependence, strengthening ties between criminal groups and the communities involved, who find mutual interest in these illicit partnerships.

Economic precariousness pushes many communities to engage in illicit activities as a means of survival. Young people, particularly those living in rural areas, are the most vulnerable, often recruited due to the lack of legitimate economic opportunities.

Social factors also play a significant role: rivalries between community leaders, local conflicts, corruption, economic hardship as well as feelings of exclusion and marginalization, weaken the social

capital, erode governance structures and undermine local authorities. In some regions, community pressure or social norms encourage involvement in criminal activities such as traditional artisanal gold mining, which is perceived as a means of collective survival. Traditional leaders may also be called upon to collect funds and provide support to communities. In areas controlled by armed and/or terrorist groups, local communities are often forced to collaborate under threat or in exchange for protection.

Resilience factors

Awareness and understanding of the interconnections between illicit markets, organized crime and terrorism contribute to strengthening community resilience in the face of these threats. Activities such as artisanal gold mining and smuggling benefit from a certain level of local legitimacy, while crimes such as cattle theft or kidnappings for ransom are perceived much more negatively. This distinction in local perceptions can influence the level of community tolerance, complicity, involvement or, on the contrary, resistance and resilience to illicit activities.

Areas with strong social cohesion have managed to limit complicity and reduce participation in illicit activities. In Bourem Cercle, for example, a strong social fabric based on trust, cooperation and mechanisms of community solidarity has helped limit the influence of criminal and terrorist groups.

Economic alternatives are a crucial element in strengthening resilience to illicit activities, as demonstrated in Sangha by a financing project providing monthly support to each household.

Basic infrastructure – such as electricity, road networks, schools and water supply – plays a central role by facilitating security interventions, strengthening social cohesion and reducing the economic and social vulnerabilities exploited by criminal networks.

The deployment of armed forces to combat violence, restore order and apprehend criminals also plays a decisive role in strengthening community resilience. Military intervention along the road between Gao and Niamey, for example, has significantly reduced crime and restored a climate of security.

> **“Inclusive, transparent local governance structures capable of responding to population needs also help reduce the appeal of illicit economies and criminal networks.”**

In Sangha, for example, a social cohesion consultation framework – bringing together the mayor, village chiefs, religious and community leaders, as well as women and youth – organized public debates broadcast live on local radio stations with the aim of countering terrorist propaganda designed to divide communities and fuel tensions.

Self-defence groups have become a key form of community resilience in certain areas such as Bandiagara. Their knowledge of the terrain and local dynamics enables them to monitor suspicious movements and contribute to reducing certain forms of crime, such as cattle theft and attacks on villages. However, their proliferation also poses challenges, including risks of abuse of power, human rights violations and the exacerbation of intercommunal tensions.

Strategic recommendations

1 | Security and infrastructure

Restore protection for vulnerable populations, particularly in the north, and strengthen essential infrastructure (health, education, water). In the south, secure urban areas and prevent the expansion of criminal networks.

2 | Countering corruption

Strengthen transparency and the rule of law to restore citizens’ trust, particularly in rural areas, through citizen audits and awareness-raising campaigns.

3 | Strengthening community networks

Mobilize communities to monitor, raise awareness and combat organized crime and terrorism, promoting social cohesion and cooperative security.

4 | Coordination of stakeholders

Improve cooperation among non-governmental organizations, the State and local structures to strengthen resilience against illicit markets and terrorism.

5 | Local and traditional governance

Support local leaders in conflict prevention and resolution, using a human rights-based and gender-sensitive approach.

6 | Improving state-citizen relations

Build trust through transparent communication and mutual accountability mechanisms, notably through spaces for citizen dialogue.

7 | Community protection

Support local councils and networks in conflict prevention and the sharing of security information, particularly through municipal security advisory committees (comités consultatifs de sécurité communaux).

8 | Regulation of non-state security actors

Regulate local defence groups to limit abuses and ensure their legitimate contribution to security.

9 | Strengthening the role of women

Promote women's role in community resilience and their involvement in prevention and awareness-raising efforts against organized crime.

Acknowledgements

The United Nations Office on Drugs and Crime (UNODC) and the United Nations Interregional Crime and Justice Research Institute (UNICRI) thank The Secretary-General's Peace and Security Sub-Fund, United Nations Peace and Development Trust Fund (UNPDF) for its generous support to this project. They are also deeply grateful to civil society representatives, practitioners and researchers whose experiences and insights informed this publication.

This report was written by Mr. Adama Mamadou Ballo, Mr. Ibrahim Ahmadou Dicko and Mr. Ibrahim Traore, national consultants of UNODC, under the overall direction of Ms. Elena Dal Santo, Programme Officer at UNICRI, Mr. Mohamed Fouda, Programme Officer at UNODC and Ms. Agathe Hazelart, Counter-Terrorism Specialist at UNODC.

The authors would particularly like to thank their local and national partners, as well as all stakeholders who participated in the consultations, meetings and workshops that led to the development of this report.

The opinions, findings, conclusions and recommendations expressed herein do not necessarily reflect the views of UNICRI, UNODC or any other national, regional or international entity involved or mentioned. Responsibility for the opinions expressed in signed articles, websites, studies and other contributions rests solely with the authors and this publication does not constitute an endorsement by UNICRI or UNODC of the views expressed herein.

DOWNLOAD HERE

THE SOCIAL RE-USE OF SEIZED AND CONFISCATED ASSETS: GOOD POLICIES AND PRACTICES

Organized crime and corruption undermine state revenues, weaken institutional credibility, and obstruct long-term development goals. Asset recovery, and in particular the social re-use of seized and confiscated assets, is a key response to ensure that illicitly-acquired resources - from financial proceeds to real estate - are not only recovered, but also redirected to serve community needs. This publication was funded by the European Union through DG ENEST under the project “Support to Eastern Partnership Countries to Enhance Asset Recovery II”.

ANALYSIS OF INTERCONNECTED CLIMATE SECURITY AND VIOLENT EXTREMISM RISKS: A PRACTICAL GUIDE FOR MAURITANIA

UNICRI’s latest methodology, «Analysis of Interconnected Climate Security and Violent Extremism Risks: A Practical Guide for Mauritania,» was developed with the generous financial support of the German Federal Foreign Office. This practical and gender-responsive tool provides a structured framework for analysing the complex intersections of climate change and violent extremism risks in Mauritania, enabling more informed, inclusive, and resilient responses.

DOWNLOAD HERE

DOWNLOAD HERE

GREEN PRISONS: A GUIDE TO CREATING ENVIRONMENTALLY SUSTAINABLE PRISONS

The impacts of climate change and environmental instability can intensify existing vulnerabilities in prisons, creating unique risks to the health and safety of those living and working within them. While prisons have a significant environmental footprint, they are often overlooked in broader sustainability discussions. At the same time, all stakeholders, including people detained in prison, can become agents of change, fostering more sustainable practices.

“As UN Women reports, 73% of women have been victims of some form of online violence, prompting 9 in 10 to reduce their online engagement

Gender dimensions and youth engagement in cybersecurity

by **Avnita Singh**

Cybersecurity has become the arena where our rights, economies, and identities are increasingly defended or lost. Yet, women and youth, two of the most important actors in shaping this domain, remain underrepresented in decision-making and overrepresented among victims of online threats. According to the International Information System Security Certification Consortium (ISC)[1] women constitute only 24% of the global cybersecurity workforce, while the most digitally active generation — those under 30 — is often sidelined from influential policy roles. This article investigates the gendered dynamics of cyber victimisation, the institutional barriers to career access, and the dual reality of youth engagement — vibrant with creativity yet persistently at risk. Drawing on case studies from Africa, Asia, Europe, and Latin America, the analysis highlights how technology-facilitated gender-based violence (TFGBV) and online radicalisation exploit existing inequalities. It addresses policy shortcomings and outlines pragmatic steps, ranging from gender conscious national cyber policies to inclusive, youth-driven global cooperation platforms. Put simply, excluding half of the population and the most dynamic young minds weakens cybersecurity preparedness.

Introduction

In 2024 alone, the world recorded over 493 million ransomware attacks, a 42% increase from the previous year.[2] Beneath these statistics lie real human impacts, which are far more evenly distributed. Women and girls face higher risks of online harassment, while youth often lack institutional support to engage meaningfully in cyber governance.

> **“A security system that excludes its own innovators is neither complete nor resilient.”**

[1] International Information System Security Certification Consortium (ISC), Cybersecurity Workforce Study, 2023

[2] SonicWall Cyber Threat Report, 2024.

Gendered threats and impact

The evidence is clear:

- **Online abuse:** As UN Women reports, 73% of women[3] have been victims of some form of online violence, prompting 9 in 10 to reduce their online engagement.
- **Deepfake exploitation:** More than 96% of all deepfake content[4] online consist of non-consensual content targeting women.
- **Economic exclusion:** The global gender pay gap in cybersecurity is 16%[5], increasing to more than 25% in lower-income regions.

Ethnicity, socioeconomic status, and disability intersect to amplify these threats, worsening their impacts and further reducing reporting.

Workforce representation and barriers

While the global cybersecurity workforce reached 5.5 million in 2023,[6] women's representation remains limited to approximately one-quarter. Even in the regions with strong Science, Technology, Engineering and Mathematics (STEM) pipelines, women are often confined to lower-paying, less influential roles. Barriers include:

- **Stereotypes:** Perceptions of cybersecurity as a male-centric profession discourage early educational engagement, narrowing future talent pipelines.
- **Networking gaps:** In cybersecurity, women are 28% less likely to have access to professional mentorship.
- **Retention challenges:**

> **"An ISC[7] survey found that 45% of women in the cybersecurity field are considering leaving their jobs within five years due to hostile work environments."**

The gender and generational dimensions of cybersecurity

The gender and generational dimensions of cybersecurity should not be understood in isolation, as they intersect in ways that compound both vulnerability and exclusion. This intersection is particularly evident in the experiences of young women and girls, who are among the most active users of digital technologies while also being disproportionately exposed to online harms such as technology-facilitated gender-based violence (TFGBV), cyber harassment, and non-consensual digital exploitation.

These vulnerabilities are shaped by broader structural inequalities, including disparities in access to education, digital literacy, and economic opportunities. Young women often face barriers to entering cybersecurity education and professional pathways due to persistent gender stereotypes, limited mentorship opportunities, and restricted access to professional networks. As a result, their

3 UN Women, Online and ICT-facilitated Violence Against Women and Girls. 2021.
4 Sensity AI, The State of Deepfakes, 2023
5 Frost & Sullivan, Women in Cybersecurity, 2022
6 International Information System Security Certification Consortium (ISC), Cybersecurity Workforce Study, 2023.
7 *Ibid.*

representation remains limited not only in the workforce but also in decision-making processes within cybersecurity governance.

At the same time, their lived experiences in navigating digital platforms provide them with valuable insights into emerging risks and online behavioural patterns. This positions them as important contributors to cybersecurity innovation and policy development.

Recognizing this overlap is essential. Without an integrated approach that considers both gender and age, cybersecurity policies risk overlooking those who are both most affected by cyber threats and best placed to contribute to their prevention.

Youth: creative innovators in the digital sphere

Youth, defined here as individuals aged 15–29, form over 50% of the online global population. Their contributions are tangible:

- University-led cyber clubs in **Kenya** have developed phishing detection tools adopted by local banks.
- Youth activists have developed WhatsApp-based misinformation scanning systems during elections in **Brazil.**
- In **India,** young trainers have organized coding bootcamps and delivered cyber hygiene education to more than 80,000 rural women.[8]

[8] International Telecommunication Union, Digital Skills Development Initiatives Report, 2023.

These cases demonstrate that when given access to tools, resources, and a platform, youth can address policy gaps and technological demands more swiftly than many institutional actors.

The dark side of digital youth engagement

> **"Yet, the same skills that make youth innovative and creative also make them vulnerable to exploitation."**

- **Cybercrime recruitment:** Europol has observed a surge in "script kiddie" recruitment, with criminal networks enticing teenagers into paid hacking activities.
- **Extremist content dissemination:** A report by the UN Counter-Terrorism Committee shows that since 2020, there has been a 300% rise in youth-targeted extremist propaganda online.
- **Mental health:** Studies indicate that excessive online exposure results in higher rates of anxiety and depression, especially when combined with cyberbullying.

Recommendations

1. **Mainstreaming gender equity:** Ensure the inclusion of gender impact assessments in every national cybersecurity policy.
2. **Youth policy inclusion:** Create national and regional youth councils focused on cybersecurity.
3. **Capacity building:** Fund targeted technical training for women and youth in underserved regions, ensuring direct transition into sustainable employment.
4. **Digital literacy:** Embed comprehensive cyber hygiene and digital rights literacy into secondary school curricula nationwide.
5. **Accountability structures:** Mandate the publication of abuse and incident reports disaggregated by gender and age across all digital platforms.

Policy Context

> **"Existing international measures offer clear entry points for gender-sensitive cyber governance."**

The Budapest Convention on Cybercrime and its Protocols can be extended to include explicit gender-responsive implementation measures. The African Union's Convention on Cybersecurity and Personal Data Protection emphasizes the need for comprehensive capacity-building initiatives. Similarly, the UN General Assembly's Open-Ended Working Group (OEWG) on information and communication technologies (ICTs) has acknowledged the importance of multi-stakeholder engagement, integrating both youth and gender perspectives — in the development of cyber standards.

Conclusion

The international cybersecurity community cannot remain a conversation among the same constrained demographic. Women and youth are not "diversity of viewpoints"— they are critical to shaping security in an interconnected world. Excluding them weakens both the resilience and the legitimacy of our cyber policies.

About the Author

Avnita Singh is a freelance writer focusing on issues related to women, social inequality, and gender bias. Her work examines the challenges emerging within contemporary youth culture in relation to digital spaces and evolving social dynamics. She mainly produces content that explores these intersections, with an emphasis on awareness, critical analysis, and social impact.

“Let safety become a right, not a reward. Let cybersecurity be justice in action. Let no one arrive online barefoot and alone again

Cyber Sakhi: a digital safety friend for those left behind

by **Maanya Chauhan**

Some step into the digital world guided by maps, guarded by laws, and fluent in its signals. Others arrive barefoot, in the dark with no instructions, no warnings. Just unfamiliar links, unfamiliar threats, and silence when something goes wrong.

Across the world's widening digital divide, millions of women and young people in underserved communities are stepping online not with curiosity, but with caution and sometimes, fear. For them, the internet is not a space of empowerment, but a maze of risk, written in a language they were never taught to read.

Cyber Sakhi[1] emerged in response to that silence. It is not just a tool or an initiative. It is a hand extended across the gap. A youth-led movement built on listening before acting, translating digital safety into languages, contexts, and care systems that mainstream frameworks overlook.

Because true cybersecurity cannot exist in isolation. It must include the invisible. It must protect the excluded. And it must begin not with code, but with compassion.

Cybersecurity, for many, is discussed in the language of firewalls, encryption, and global policy. But in underserved communities, the threats are closer to home and far more personal: a message that steals a widow's savings, a stranger who hijacks a girl's identity, or an app that watches silently, without consent.

> **"In places where access to technology is growing faster than access to education, millions are being left digitally exposed."**

[1] Cyber Sakhi is an independent, student-led cybersecurity initiative that I founded, focused on improving digital safety awareness among underserved communities in India. Through hands-on workshops and a lightweight Chrome extension, it supports first-time internet users – especially women and youth – in recognising online risks through simple, multilingual, and easy-to-understand guidance.

For first-time internet users, especially women and youth in rural India and beyond, the internet arrives not as a gateway to progress, but as a battlefield without armour.

Today's threats evolve faster than defences. AI-generated scams mimic trusted voices. Spyware hides in ordinary apps. In local dialects, misinformation spreads unchecked. First-time users face an invisible war that no one has taught them the rules.

Too often, workshops assume prior knowledge.

> **"Tools are created by those who have never sat in a community hall with a broken fan and no Wi-Fi. And policies are written far from the voices they claim to protect."**

What is missing is not awareness, it is access. Access to safety that feels familiar. Solutions that feel human. And digital empowerment that does not shame or intimidate, but understands.

This is not just a gap in technology. It's a gap in empathy. One that continues to widen as global digital inclusion accelerates without guardrails for the most vulnerable.

And it's not unique to India. Across underserved communities globally, young users are stepping online into the same darkness, with no tools and no shields. This is a shared global silence, and we must respond together.

Cyber Sakhi did not start in labs or boardrooms. It emerged from whispered concerns and quiet stories shared behind workshop doors, where a single wrong click carries more fear than any promise the internet holds.

We began with listening. Not surveys or checklists, but real conversations with women, girls, and youth who had been left to fend for themselves online. What we heard shaped everything: the confusion, the isolation, the quiet aftermath of harm.

So, we responded with something simple but radical: care-informed cybersecurity.
Cyber Sakhi delivers hands-on digital safety workshops in local languages, in familiar places —

schools, NGOs, and even under trees when walls are not available. We demystify the web, decode scams and explain settings that protect instead of confuse. Each session builds a bridge from fear to agency.

To extend our impact beyond the classroom, we built a lightweight Chrome extension tailored for first-time users. It flags unsafe links, alerts users in real-time, and provides friendly prompts in simple language. Just support, exactly when and where it's needed.

But Cyber Sakhi is not just a service, it is solidarity. It is tech that remembers the human. It is a promise that no one should face the internet alone.
What does change look like when it's not captured in spreadsheets?
It looks like a grandmother in a village using a borrowed smartphone to report a scam message, and then teaching her neighbours how to spot one. It looks like a girl, once silenced by online harassment, now standing in front of her class explaining two-factor authentication.

Cybersecurity is no longer only about firewalls and passwords. It is a question of justice. Of dignity. Of who gets to feel safe online, and who is left to fend for themselves.

> **"The digital divide is no longer just about access. It is about agency. As technology evolves, so must our empathy, our systems, and our solutions."**

Tools alone rarely heal the trauma of being silenced or scammed. It takes people, community, and a commitment to showing up where no one else does. Cyber Sakhi is more than a project. It is a beginning.

It is grassroots innovation rooted in care. It is digital resilience built from the bottom up. It is proof that when young people lead with empathy, they can build what policies overlook and power structures delay.

And if one young changemaker can do this with limited resources, imagine what becomes possible when institutions, innovators, and communities move in sync in partnership.

The future of cybersecurity is not written in code. It is written in compassion.

> **"The question is not whether we can protect the most vulnerable. It is whether we choose to."**

And if we do, then safety will no longer be a privilege. It will be a promise.

Yet, this is only the beginning. Cyber Sakhi is a seed — not a solution in full bloom, but proof that empathy, when paired with technology, can shift realities. We dream not of scaling metrics, but of expanding dignity. Of a world where every woman, every girl, every first-time user logs on without fear.

We imagine local champions leading safety circles in their villages. Policymakers who listen to lived experiences before writing laws. Global institutions resourcing not just the largest organizations, but the most courageous grassroots initiatives.

The Internet should not belong only to those who are digitally literate, powerful, or already protected. Its future must be shaped by those who were once silenced and who now speak, not only in technical terms, but also in the language of care.

If Cyber Sakhi can emerge from whispered fears and still find a voice, then what else becomes possible when the world starts listening?

Let safety become a right, not a reward. Let cybersecurity be justice in action. Let no one arrive online barefoot and alone again.

About the Author

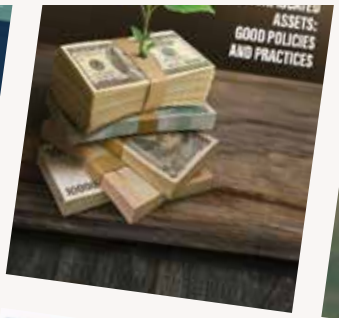
Maanya Chauhan is an 18-year-old cybersecurity undergraduate and youth changemaker from India, passionate about bridging the digital safety gap in underserved communities. She is the founder of Cyber Sakhi, an initiative that empowers first-time internet users, especially women and youth, through hands-on workshops and a custom-built Chrome extension offering real-time privacy alerts. Maanya has led digital rights sessions in collaboration with grassroots NGOs and is committed to building inclusive, human-centered technology that centers empathy, justice, and access for all.

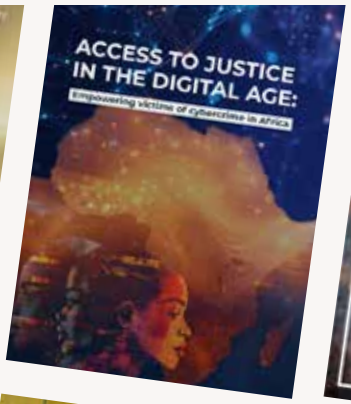
→ **The images featured in this article are courtesy of the author and are drawn from the Cyber Sakhi programme.**

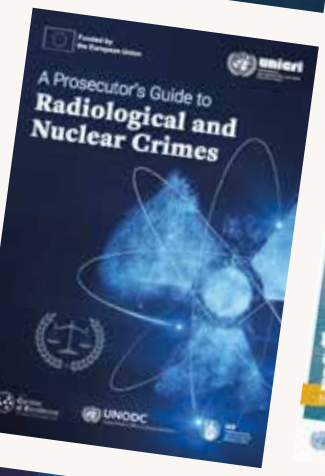
unicri
United Nations
Interregional Crime and Justice
Research Institute

Publications

Download UNICRI publications







Beyond the binary: empowering youth as agents of change in cybersecurity and crime prevention

by **Per-Albin Johansson**

Global security and justice face a striking paradox. Despite youth crime's complex and profound societal repercussions,

> **"young people under the age of 24 (approximately 16% of the world's population[1]) remain largely excluded from decisions that will significantly shape their future."**

This disparity highlights a crucial need to shift the current approach from merely addressing youth crime as a challenge to actively engaging young people as key partners in reshaping crime prevention strategies.

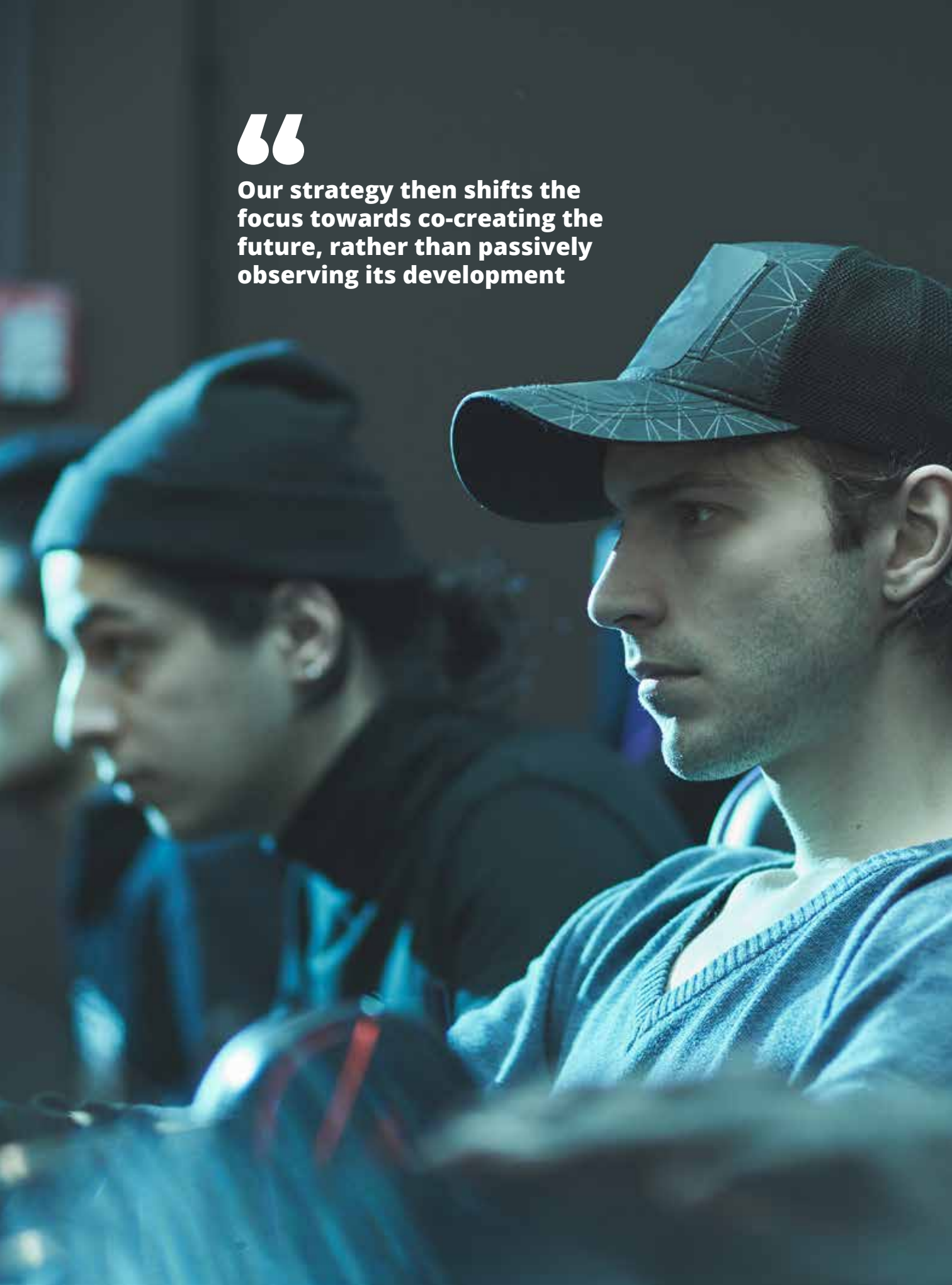
Simultaneously, it is evident that digital environments are emerging as one of the most prominent arenas of contemporary crime. This phenomenon extends far beyond traditional notions of offline criminal activity, encompassing a rapidly evolving spectrum of offences that occur online or are significantly facilitated by the Internet and related digital technologies. We are witnessing a fundamental reshaping of the criminal landscape, whereby cyber threats are not merely a specialized category but an increasingly pervasive element of modern criminality.

These 'digital and digitally mediated crimes' can broadly be categorized as cyber-dependent or cyber-enabled.[2] The first refers to offences that can only be committed using information and communication technologies (ICTs) or within the digital realm.

However — and more importantly concerning youth and adolescents — the second category comprises traditional crimes that are amplified, extended, or even made possible through the use of digital technologies. This category includes fraud, child sexual exploitation, illegal trade in

[1] United Nations, *World Youth Report 2020* (New York: Department of Economic and Social Affairs, 2020).

[2] European Parliamentary Research Service, *Cybercrime and Cybersecurity*, Briefing 760356 (Brussels: European Parliament, 2024).



Our strategy then shifts the focus towards co-creating the future, rather than passively observing its development

goods (drugs, weapons, etc.), and perhaps most critically, criminal recruitment.[3]

The increasingly blurred lines between the virtual and physical worlds contribute to the stark reality of escalating youth recruitment into criminal networks, a challenge that is particularly pressing in the Nordics.[4]

> **"Popular social and gaming-adjacent platforms, such as Roblox and TikTok, are increasingly used by criminal actors for recruitment purposes."**

The ease of access and anonymity these platforms afford allows recruiters to target and manipulate vulnerable adolescents. Through seemingly innocuous interactions, young people are subtly drawn into illicit activities. After instrumentalizing the individual for their purposes, those who orchestrated the initial crime disappear, leaving the new 'perpetrator' to carry out the illicit activities and, more importantly, face the consequences alone.

In parallel, the predominantly analogue legal and policy systems frequently lag behind arenas where crime increasingly risks creating lawless spaces: mainly in cyberspace.[5]

This evolving landscape reveals the profound inadequacy of many traditional legal constructs, particularly the foundational victim-perpetrator binary, to address modern criminal dynamics. Adolescents often oscillate between being targeted and being coerced into committing offences; consequently,

[3] Johansson, Per-Albin, "Northern Sprites: Gamifcation and Youth Recruitment in the Nordic Region, "*GNET Research*, 30 October 2024.

[4] Tollin, Katharina, Angerbrandt, Henrik and Jonsson, Anna, *Children and Youth in Criminal Networks: Network Entry, Offending, Conditions, and Network Exit,* Report 2023:13 (Stockholm: Swedish National Council for Crime Prevention, 2023).

[5] Europol, *Policing in an Online World* (The Hague: Europol, 2025).

> **"the lines between being an impressionable individual groomed and subsequently compelled into illicit acts are as blurred as those between the virtual and physical realities."**

Without clear and legitimate conceptualization, many young people are left in a legal and ethical limbo, where traditional definitions of culpability and victimhood offer little clarity or protection.

Youth caught up in episodes of violence, such as those observed in Sweden[6], are often placed in a precarious position and reductively categorized as either passive recipients of harm or singularly responsible perpetrators. This problematic framing highlights a critical inadequacy that is not merely a fallacy, but an opportunity to rethink our approach fundamentally. Rather than maintaining this dichotomy, we must champion a transformative method that recognizes and cultivates their potential as active agents of change before they become entangled in the cycle. Beyond their aforementioned and acknowledged vulnerabilities to cyber-dependent or cyber-enabled crimes, youth represent a powerful, often untapped, resource.[7]

Observations from contexts such as Sweden, as well as the Netherlands, indicate that the critical challenge in empowering youth as agents of change often lies not in their disinterest but in outdated communication strategies.[8]Adults frequently talk about young people, rather than with them, employing language, formats, and channels that fail to resonate. This oversight perpetuates the misconception that digital security is an inherently technical or frightening domain, leading to immediate disengagement. To truly connect with young

6 Europol, *Policing in an Online World* (The Hague: Europol, 2025).

7 United Nations Offce on Drugs and Crime (UNODC), Crime Prevention and Youth (Vienna: UNODC, n.d.).

8 Schiks, J.A.M., van 't Hoff-de Goede, Susanne and Leukfeldt, Rutger E., "An Alternative Intervention for Juvenile Hackers? A Qualitative Evaluation of the Hack_Right Intervention," Journal of Crime and Justice, vol. 47, no. 4 (2024): 492–510.

audiences, the discourse around digital safety must be reframed to align with their lived digital realities: encompassing social media interactions, online gaming, gamified environments, and, especially, peer relationships. This approach should therefore focus on demonstrating how proactive digital safety directly impacts their everyday lives. Practitioners should recognize that peer influence is a potent force among adolescents, accordingly, a powerful strategy emerges: cultivating a movement of digital role models on the very social and gaming-adjacent platforms where they spend much of their time.

This methodology is effectively realised by public-private partnerships like HackShield Future Cyber Heroes and its extensions, which engage youth through game mechanics in simulated and safe digital environments. Crucially, such an approach fosters their active agency.

> **"Selected participants, collaborating directly with both law enforcement and cybersecurity specialists, move beyond conventional learning to actively share their knowledge.[9]"**

Both in Sweden and the Netherlands, they act as intermediaries by conducting presentations for their classmates, leveraging peer influence to illustrate digital threats in a simple manner. Moreover, these young agents have gone on to facilitate crucial intergenerational knowledge transfer by subsequently advising seniors on online safety practices.

By supporting these youth-led voices through similarly integrated programs, it is possible to foster engagement and create a more relevant, accessible, and ultimately effective culture of cybersecurity awareness and resilience.

> **"The malevolent creativity of cybercrime necessitates equally creative responses, and neglecting the inherent agency of young people, with their innate understanding of online environments, is a critical gap that must be addressed."**

True progress hinges on recognizing youth not as passive subjects or singular perpetrators, but as indispensable agents of change. This engagement empowers them to not only protect themselves and their peers but also to defend acquaintances — such as vulnerable family members targeted by fraud or manipulation — from the growing impact of organized crime in digital environments. In essence, our strategy then shifts the focus towards co-creating the future, rather than passively observing its development.

[9] Spithoven, R., Leukfeldt, R., Misana-ter Huurne, E., van 't Hoff-de Goede, S., van Houten, Y., Bekkers, L., Foppen, E. and te Bos, J., *Cyberweerbaarheid: Een gemeentelijk offensief ter preventie van slachtofferschap van cybercrime* (The Hague: n.p., 2022).

About the Author

Per-Albin Johansson is an experienced conflict researcher and specialist in digital crime prevention who bridges the gap between complex academic theory and tangible, everyday security solutions. As the Safer Sweden Foundation's main expert on cyber-enabled crime, he utilises a multidisciplinary background in crisis management and conflict transformation to interpret the modern threat landscape. Whether researching online vulnerabilities or implementing agency-centred initiatives, Per-Albin is driven by the goal of eliminating the offender's opportunities for crime rather than restricting the victim's fundamental right to digital participation.

Despite such threats, rural youth are transforming from passive users into frontline cyber defenders

From village Wi-Fi to virtual battlefields: how rural youth are becoming cybersecurity's frontline

by **Santhos Sivan**

As Internet access expands, especially in rural India, cybercrime has surged by over 400% between 2021 and 2024, with villages and small towns becoming new hotspots — underscoring an urgent need for grassroots resilience. Despite such threats, rural youth are transforming from passive users into frontline cyber defenders. This article explores how they self-train through low-cost innovations — like do-it-yourself (DIY) antenna-based Wi-Fi extensions and community digital safety clubs — to protect local networks and combat rising attacks in the absence of formal resources or recognition. It references a landmark Amroha Police cybersecurity training programme, which empowered over 500 students from 22 Indian states with skills in ethical hacking, digital forensics, open-source-intelligence (OSINT), and Capture-the-Flag (CTF) exercises. Addressing structural limitations — patchy infrastructure, limited digital literacy, and policy neglect — it calls for targeted investment, recognition, and rural inclusion in national cybersecurity strategies.

> **"These rural cyber pioneers prove that the most resilient defences can emerge not from high-tech labs but from resourceful communities under banyan trees."**

Empowering them is not only a matter of equity — it is essential for global cyber safety.

Necessity meets opportunity

Smart infrastructure initiatives like Project Bharat-Net have connected over 213,000 Gram Panchayats by August 2024, expanding rural access — but often without accompanying cybersecurity safeguards. As a result, local endpoints — internet cafés, Common-Service- Centre (CSC) kiosks, and outdated government computers —

have become entry points for malware that harvests sensitive data like Aadhaar information or subsidy details.

In this fraught landscape, self-taught young defenders have emerged. Under banyan trees and in modest common rooms, they set up digital safety clubs and teach their communities to spot scams and secure devices.

> **"In one Tamil Nadu village, teenagers built DIY Wi-Fi extensions from discarded routers — bringing secure connectivity to entire communities."**

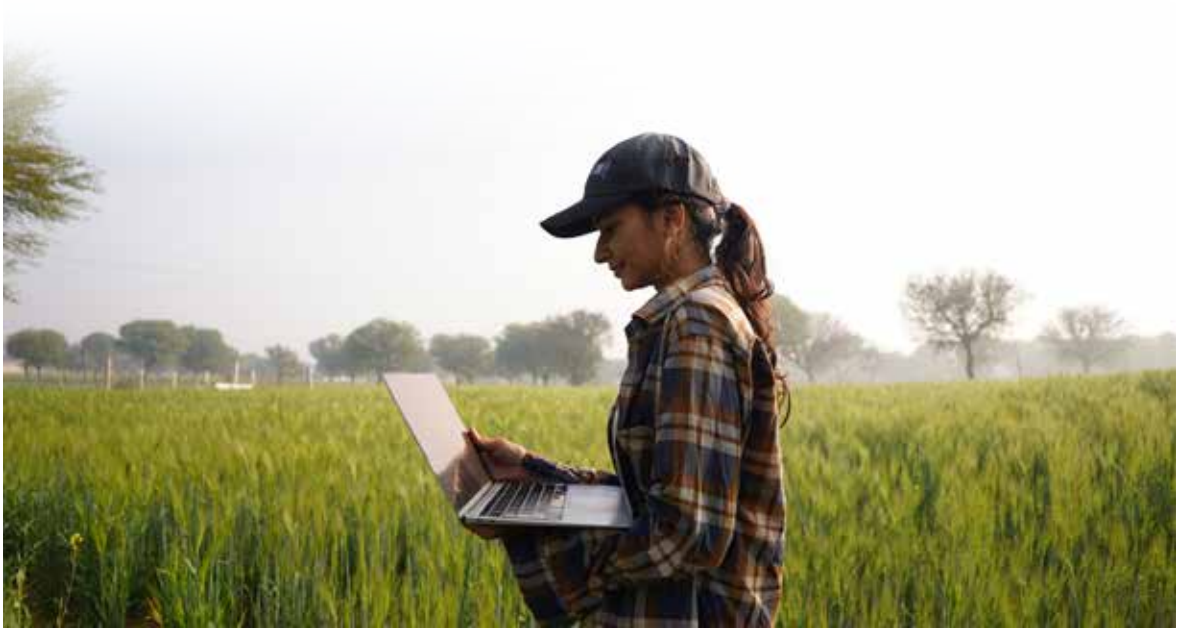
In Kenya, youth-developed apps that function without internet access in local dialects helped elders recognize phishing attempts.

Case study: institutional engagement

In June 2025, the Amroha Police in Uttar Pradesh launched a pioneering cybersecurity internship program — training over 500 students from 22 states in digital forensics, ethical hacking, the dark web, and OSINT via hands-on Capture-the-Flag events and case study simulations. Participants, including many from rural Tamil Nadu, have begun outreach in their villages — highlighting the power of localized, peer-led training models.

Structural challenges persist

Despite rising Internet use — 82% of rural youth (15–24 years) can now access the web — only 26.8% engage in advanced online activities like emailing and banking. Cyber literacy remains low: over 90% of Indians lack basic digital safety awareness. Rural Small and Medium-sized Businesses



(SMBs) and Micro, Small and Medium-sized Enterprises (MSMEs), key pillars of the economy, often operate without any cybersecurity measures, leaving them vulnerable to information-stealing malware and fraud.

Global and cross-layer relevance

The DEF CON Franklin initiative in the United States offers a powerful parallel — deploying volunteer hackers to secure rural water systems through no-cost upgrades and network assessments. Similarly, cyber-attacks on rural water utilities in Texas reportedly involving foreign threat actors, caused system overflows — averted only by manual intervention. These cases underscore how rural vulnerabilities can ripple into broader infrastructure crises — and how locally trained, community-based defenders, locally trained, can be pivotal.

Three steps to strengthen rural cyber resilience

1. **Localized, peer-focused training:** – Scale programs like Amroha's, CyberVaahini, and CyberPeace Foundation's e-Saksham to build a network of rural youth trainers who understand local languages, contexts, and constraints.

2. **Micro-grants and resource kits:** Fund small but critical tools — DIY router kits, trainers and offline learning materials. A micro-grant could enable a village youth group to shield hundreds from malware or scam networks.

3. **Policy inclusion and recognition:** Integrate rural cyber initiatives into national strategy. Formalize Young Defender programs through the Indian Cybercrime Coordination Centre (I4C) and the Indian Computer Emergency Response Team (CERT-In) and recognize community defenders as vital stakeholders in India's cybersecurity architecture.

Conclusion: redefining cybersecurity frontlines

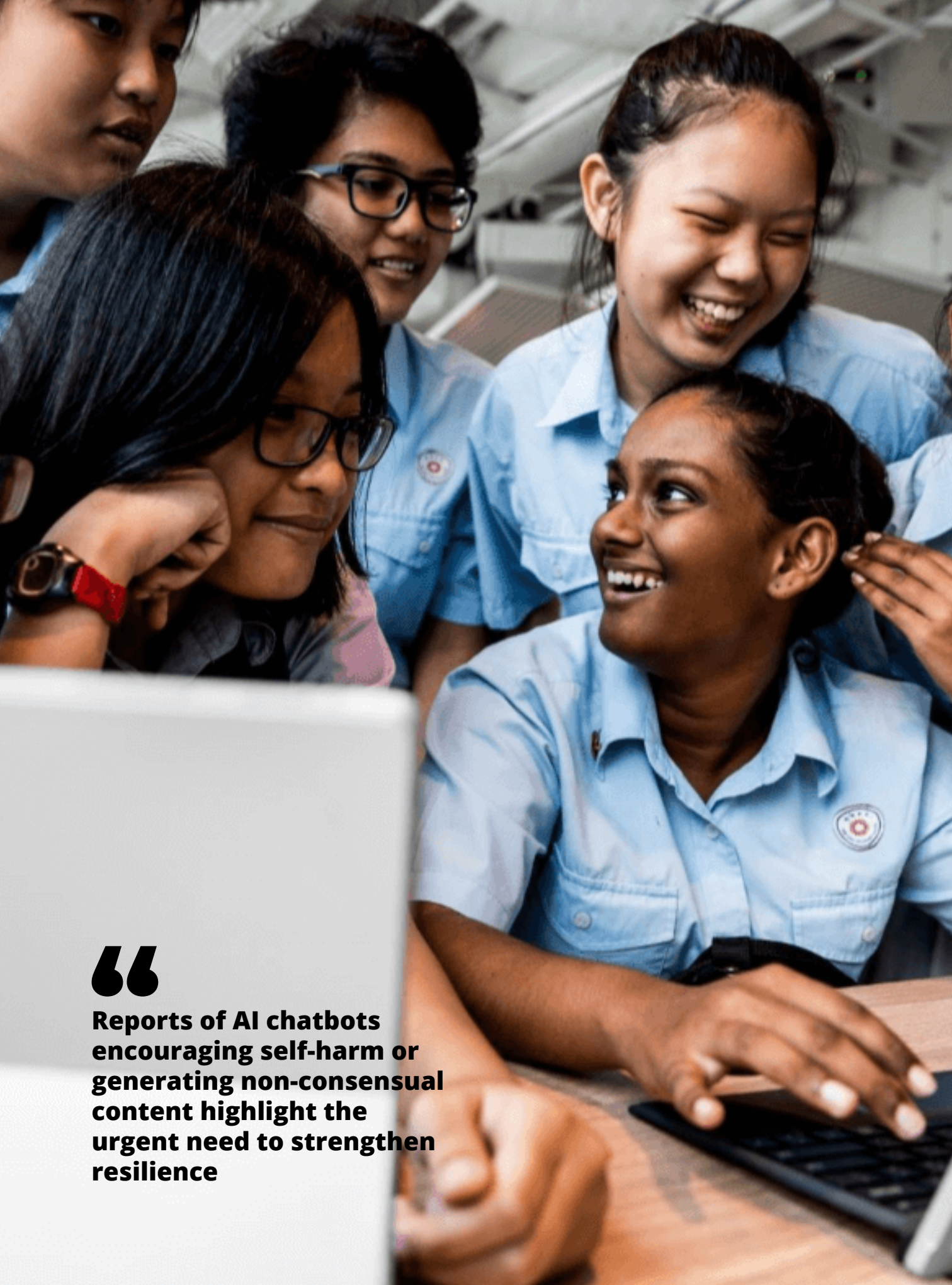
The surge of cyber threats is not just a technological crisis — it is a societal one. Rural youth, armed with innovation, solidarity, and a sense of stewardship, are often the first responders in their communities.

> **"To build a truly resilient digital future, we must expand beyond the urban lens and empower these underrepresented defenders."**

The strongest firewalls may well begin with grassroots movements — demonstrating that digital safety is built one Wi-Fi signal, one trained young defender, and one trusted community act at a time.

About the Author

Santhos Sivan is a cybersecurity enthusiast, social impact writer, and auditor from Tamil Nadu, India. With a background in cooperative systems and a keen interest in technology's role in justice, he explores how underrepresented communities can shape global security narratives. His work bridges grassroots realities with policy-level discussions, advocating for inclusive and ethical digital futures. Passionate about empowering rural youth, he documents how innovation often emerges in unexpected places. Santhos's writing combines analytical depth with storytelling, aiming to foster dialogue, challenge assumptions, and inspire actionable change in cybersecurity, governance, and human rights.



Reports of AI chatbots encouraging self-harm or generating non-consensual content highlight the urgent need to strengthen resilience

Digital guardians of the AI era: building youth cybersecurity resilience

by **Ziarla Mae Malabanan**

Artificial intelligence (AI) is reshaping how young people communicate, learn and express themselves online. At the same time, it exposes them to unprecedented risks – from algorithmic bias and privacy violations to harassment, exploitation and AI-driven scams. These threats are not hypothetical: reports of AI chatbots encouraging self-harm or generating non-consensual content highlight the urgent need to strengthen resilience.

This article argues that youth-led cybersecurity capacity building is an effective defence against AI-enabled threats.

> **"By combining technical training with policy literacy, ethical awareness and hands-on experience, young people can become proactive digital guardians able to detect, respond to and prevent cyber harm."**

Drawing on case studies from Southeast Asia, Europe and Africa, the article demonstrates how inclusive programmes, particularly those targeting underrepresented or vulnerable communities, empower youth to navigate digital risks confidently and responsibly. Investing in youth resilience not only strengthens individual safety but also creates a new generation capable of strengthening digital ecosystems.

Introduction

Young people are increasingly at risk from artificial intelligence, the United Nations Children's Fund[1] and the United Nations Educational, Scientific and Cultural Organization[2] warn. The risks include algorithmic bias, privacy violations and exposure to manipulated content. Tools like chatbots, AI-driven teaching applications, and content generators are transforming learning, social connection, and self-expression. While these tools can entertain or educate, they are often used without

[1] United Nations Children's Fund (UNICEF), Generative AI: Risks and Opportunities for Children (Florence: UNICEF Innocenti, 2023)

[2] UNESCO, Ethics of Artificial Intelligence – The Recommendation.

supervision, leaving youth vulnerable to data collection, manipulation, and exploitation.

An investigation by Triple J Hack, an Australian current affairs programme, revealed allegations of AI chatbots sexually harassing and encouraging self-harm of young people, as well as allegations of ChatGPT reinforcing delusions that led an individual to hospitalization. Young people connect with chatbots or AI companions (digital characters powered by AI) as a form of social outreach or emotional support, said a youth counsellor interviewed for the programme. In some cases, they were told they had "no chance of making friends", that they were "ugly or disgusting", or that they should "kill themselves".[3] Other similar incidents – including harassment during language learning – show these risks are neither isolated nor hypothetical.

"Beyond chatbots and similar tools, cybercriminals leverage AI to automate data collection, profiling, cyberbullying, and the creation of deepfake child sexual abuse material."

Even digitally fluent youth are vulnerable: LinkedIn-based employment scams, for example, steal sensitive personal or financial information by targeting college students and early-career professionals with convincing but fake job offers. Deepfake technology, once niche, is now widely accessible, as shown in a 2020 investigation, which enabled users to generate over 100,000 non-consensual images of minors generated by bots on Telegram.[4]

[3] A. McLennan, AI Chatbots Accused of Encouraging Teen Suicide as Experts Sound Alarm, ABC News, 12 August 2025.

[4] R. Karim and M. Seera, Digital Child Abuse: Deepfakes and the Rising Danger of AI-Generated Exploitation, Monash Lens, 25 February 2025.

These examples, showing that AI increases the online risks youth face, also highlight the importance of cybersecurity resilience. One of the most effective long-term solutions is youth-led capacity building in cybersecurity. By participating in or leading these programmes, young people develop digital literacy, preparedness, and resilience, enabling them to navigate AI-enabled threats safely. This type of resilience not only shields individuals but also strengthens the overall integrity of digital ecosystems.

Framing cybersecurity education and AI awareness around youth – defined as persons aged 15–24[5] – empowers digital natives to become proactive, informed, and resilient participants in the online world.

Youth as digital guardians

Because young people already live in digital ecosystems, they learn and adopt new technologies quickly. However, tech fluency alone is not enough. They also need training in policy literacy, including how laws, regulations, and data governance operate in cyberspace. This objective can be achieved by participation in campaigns, workshops, and opportunities to engage directly with policymakers.

> **"With appropriate training, young people can turn their technological fluency into a defensive advantage, spotting AI-driven manipulation or suspicious patterns more naturally than less tech-immersed groups."**

[5] United Nations, Frequently Asked Questions United Nations for Youth, United Nations.

While early exposure to cybersecurity creates a culture of safe behaviour by default, it also equips youth to shape future policies, products, and education. Early engagement in cybersecurity and AI ethics through a bottom-up approach helps reinforce safer, more responsible digital environments.

Case studies in youth-led resilience

Across the globe, non-profit groups are investing in the next generation of cybersecurity leaders.

In 2022, the Association for Southeast Asian Nations (ASEAN), in partnership with the ASEAN Foundation and Microsoft Asia Pacific, launched a regional programme to raise awareness, expand knowledge, and strengthen cybersecurity skills among young people. What distinguished this programme was its focus not on elite universities, but on underprivileged and often overlooked communities, including unemployed or underemployed youth and female-only groups breaking into information technology.

> **"Seven ASEAN countries, together with the FutureReadyASEAN.org initiative, delivered free, localized training reaching tens of thousands, proving cybersecurity can empower youth."**

A representative from CyberGuardiansPH, one of the partners of the programme, said: "The ASEAN Cybersecurity Skilling Programme has allowed the youth to appreciate cybersecurity as a tool to protect themselves online. We have participants with no IT or STEM backgrounds who realised that cybersecurity is not as intimidating as it sounds".[6] Beyond participants, even some trainers have since pursued careers in cybersecurity, a trend echoed by participants across all seven countries.

In Europe, the CyberPeace Builders, run by the Geneva-based CyberPeace Institute, mobilize over 1,300 volunteer professionals to safeguard non-governmental organizations. Since 2021, they have completed over 1,150 missions. In one example, participants helped a humanitarian organization in Eastern Europe recover from a ransomware attack, regaining access to critical donor databases and restoring operations within 48 hours.

South Africa's CyberM8 initiative trained thousands of learners, helped small information and communications technology businesses adopt advanced digital skills, and sparked a nationwide culture of cyber awareness. For example, this initiative guided a local tech start-up to implement robust cybersecurity protocols to protect against a phishing attack on their client database. In another example, CyberM8 hosted workshops that reached over 500 students in a single week. These are two examples of youth-led initiatives that have contributed to strengthening cybersecurity.

From Southeast Asia to Europe and Africa, these initiatives prove that when youth are given inclusive, hands-on training and real-world opportunities, they become the first line of defence against the world's pressing digital threats. At their core, these initiatives follow a clear strategy: They combine education, practical support, and community-focused empowerment in ways that can scale and adapt to local cultures.

[6] Z. Malabanan, "A safe cyberspace for the ASEAN community,". ASEAN Magazine, 2022,

Pathways to youth-led cyber resilience

As AI transforms daily life, technical fluency alone is insufficient to address an increasingly complex threat landscape. Young people must pair technical skills with ethical, social, and legal awareness. AI-driven threats – from deepfakes to automated phishing – outpace legislation and traditional security tools, demanding human defenders capable of adapting to this ever-changing digital reality.

Investing in youth-led capacity building achieves exactly this objective. Providing inclusive, hands-on programmes that blend technical skills with ethics, policy literacy, and community engagement, will cultivate a generation of cybersecurity-savvy youth who do more than simply react. They will innovate, develop new tools and defensive AI systems, and challenge predictable patterns of attack with creative problem-solving. This approach will ensure a sustained movement toward responsible digital users and resilient cyber ecosystems.

About the Author

Ziarla Mae Malabanan is a certified Project Management Professional building a full-time career in cybersecurity. She has led youth empowerment and capacity-building programmes across ASEAN and Europe, including leading the 2017 ASEAN Youth Engagement Summit, helping establish the Student Youth Think Tank for Europe-Asia Relations in 2021, and directing a regional ASEAN cybersecurity programme in 2022. Her work equips young people with the skills, knowledge and leadership to navigate digital risks responsibly.

JOIN OUR SUMMER SCHOOLS

Specialized programs in Migration and Human Rights, Artificial Intelligence (AI) and Ethics

Migration and Human Rights

Gain practical knowledge on global migration challenges, human rights protection, and policy development through expert-led sessions, case studies, and interactive exercises.

Artificial Intelligence, Ethics and Human Rights

Explore the intersection of artificial intelligence, ethics, and human rights through practical exercises, expert insights, and real-world case studies.

unicri
United Nations Interregional Crime and Justice Research Institute

“As cyber threats grow in scale and complexity in this rapidly evolving digital ecosystem, youth engagement is indispensable – not merely in supporting cyber security, but also in shaping it

Youth are at the heart of cyber resilience

by **Alessia Balsamo**

In recent years, the global cyber threat landscape has transformed dramatically. Early on, well-known threats such as malware, ransomware and phishing campaigns were commonplace. Today that old landscape is being changed by far more sophisticated and intelligent tools, capable of deceiving even the most skilled cybersecurity experts.

Bolstering cyber resilience, however, requires more than technical expertise and solutions. This new digital landscape demands the active engagement of everyone – in particular young people who, having grown up inside this environment, are uniquely positioned to be co-creators, educators and defenders. As cyber threats grow in scale and complexity in this rapidly evolving digital ecosystem, youth engagement is indispensable – not merely in supporting cyber security, but also in shaping it.

Companies today face significant and converging challenges: geopolitical tensions that intensify their global risk exposure; unpredictable cyber attack patterns that complicate their planning and response; and their rapid adoption of emerging technologies that create previously unknown vulnerabilities.[1]

Compounding these global challenges is a widening cybersecurity skills gap.

> **"In 2025, nearly half (49 per cent) of public-sector organizations cited a lack of skilled personnel as a barrier to meeting their security objectives"**

an increase of 33 per cent from the previous year.[2] This is not merely a perception; the data confirm it as a worsening trend: from 2020 to 2024, 56 per cent of all major cyber incidents recorded since 2011 occurred in those five years alone.

[1] World Economic Forum. Global Cybersecurity Outlook 2025. Geneva: World Economic Forum, 2025.

[2] *Ibid.*

More striking, in comparison to 2019 – prior to wide spread remote work and the emergence of pervasive artificial intelligence (AI) – major incidents have increased by 112 per cent with a monthly average that rose from 139 incidents in 2019 to 295 in 2024.[3] A shortage of skilled personnel coupled with a staggering increase in the number of cyber attacks remains an obstacle to effective cybersecurity.

Cyber attacks inflict substantial economic, legal, reputational and technological damage, and know no geographical bounds. In 2024, 65 per cent of worldwide incidents targeted Europe and the Americas, a large number that may in part be due to strong disclosure obligations under the European Union's General Data Protection Regulation, the Network and Information Security Directives, and the Digital Operational Resilience Act. Nevertheless, within that geographic area, Europe alone still recorded a 67 per cent surge. Considering Oceania, that region saw a 228 per cent increase in the number of cyber attacks, largely attributed to newly enforced cybersecurity disclosure policies.[4]

In terms of the types of attacks in 2024, cybercrime remains dominant, accounting for 86 per cent of global incidents. Cyber criminals have adopted the so-called as-a-service model – where criminal tools are rented – which has made cybercrime more profitable and accessible, especially to individuals with minimal technical skills.

"Traditional organized crime now reinvests proceeds from offline operations into cyber activities, increasing their threat capacity."

[3] Clusit. Rapporto Clusit 2025 sulla sicurezza ICT in Italia e nel mondo. Milan: Clusit, 2025

[4] *Ibid.*

Outside of cybercrime, hacktivism and information warfare nearly doubled in 2024, while espionage and sabotage declined.[5]

The social and economic impacts of cybercrime cannot be underestimated. A much-cited example is the 2024 CrowdStrike incident in which a software update crashed 8.5 million Windows systems globally, resulting in significant disruption to airports, hospitals, financial institutions, and businesses. In that cyber incident alone, companies experienced more than USD $5 billion in direct losses.[6]

As cyber threats grow in scope and complexity, there is a rising demand for new perspectives from law, psychology, communication, and other non-technical fields. These disciplines will inform the social, behavioural, and strategic dimensions of cybersecurity. Intensifying this need is workforce strain. By the end of 2025, nearly half of global cybersecurity leaders were projected to leave their roles due to burnout. Beyond that, 66 per cent of Chief Information Security Officers reported that expectations were excessive. In fact, over half of them had experienced burnout in the previous year. Workforce burnout remains a troubling concern in the present digital ecosystem.

Amid the demand for new perspectives in a diminished workface, it is not surprising that the gap in talent is widening. In fact, two in three organizations report a critical shortage of cybersecurity professionals; only 14 per cent express confidence in their current capabilities.[7]

Youth engagement presents a viable solution to current cyber challenges. Yet bridging this talent gap requires a broader, more inclusive approach to talent development – one that empowers young people from all backgrounds to contribute fully to cyber resilience.

5 Clusit. Rapporto Clusit 2025 sulla sicurezza ICT in Italia e nel mondo. Milan: Clusit, 2025
6 CNN. CrowdStrike outage: cause and cost CNN Business, 24 July 2024.
7 World Economic Forum. Global Cybersecurity Outlook 2025.

Young people grew up in digital environments; they have never known anything else. With ease, they navigate informal channels and emerging platforms both of which are increasingly targeted by cyber threats, such as identity theft, phishing, and AI-generated deepfakes.

> **“Fluency in digital culture, creative tools, and online social dynamics are the reasons why youth are well equipped to strengthen cyber resilience from the inside.”**

Yet, being embedded in this digital space, while essential, is not enough – youth engagement must extend beyond technical skills alone.

Young people today live in a hyper-connected world, which does not in itself produce the critical awareness or understanding needed to navigate this digital space. Cyber awareness remains an essential skill not to be underestimated. To help young people understand the implications of their online actions and the value of their personal data, new approaches are needed. Those actions include: digital peer education that enables youth to serve as content creators within their communities; regional learning hubs that offer AI and cybersecurity training; and mentorship in collaboration with youth organizations and global platforms that strengthen and share the knowledge of the tools used daily.[8]

Today’s cyber threat landscape is fast-moving, AI-driven, and increasingly complex, leaving no sector or region untouched. Traditional, purely technical responses, can no longer keep pace with the speed and sophistication of modern cyber attacks. So what is the solution?

Investing in early cyber awareness, fostering capacity-building before workforce entry, and opening inclusive career pathways for individuals from diverse backgrounds, young people can benefit from their status moving from passive digital consumers into proactive defenders. Closing the skills gap, countering burnout, and building resilience will require a generational shift, one that fully embraces youth as active partners in cybersecurity.

About the Author

Alessia Balsamo is a cybersecurity professional specialized in IT risk with 7 years of experience in Security for Financial Services. Her academic background includes a law degree with a focus on international criminal law from the Milano Bicocca University during which she had the opportunity to study also in Lithuania (Erasmus) and Spain (Summer School). In 2025, she represented Italy in the official Youth engagement Group of the G7 (Y7) for AI and Digital Technology and in October she began the Master of Laws (LL.M.) in Cybercrime, Cybersecurity and International Law organized by UNICRI.

[8] Youth 7. Y7 Summit Communiqué 2025. Rome: Youth 7, 2025.


“
Youth engagement presents a viable solution to current cyber challenges


“
Cyber peace
begins with us

Cyber peace in cyberspace: empowering youth for the safe and fair use of technology

by **Princess Tudtud**

Technology has evolved significantly over the past few decades and so have the threats that cause harm rather than convenience. Cybersecurity enables users to operate computers and other digital devices safely and efficiently. All of us – even you – should engage in cybersecurity practices to protect our personal information, online activities and identities. The intersection of justice, peace and gender equity is often overlooked, resulting in limited promotion of fair use and contradicting the core principles of netiquette – understood as respectful and responsible online behaviour. As the world becomes increasingly reliant on digital platforms, with transactions and interactions moving online, cybercrime, misinformation and harassment disproportionately affect vulnerable populations. This article explores how youth engagement, gender equality and ethical practices can contribute to a secure and inclusive digital environment, alongside the growing accessibility of the internet and technology worldwide.

In the 21st century, cyberspace has become a double-edged sword, making the responsible use of technology especially important for young people. It offers access to information, freedom of expression and innovations. At the same time, it exposes users to threats that are borderless, faceless and rapidly evolving – often outpacing human responses. As an actively engaged student from the Philippines, I have witnessed this firsthand, including breaches of official school systems and online platforms. This is a call to promote the responsible use of technology and to protect ourselves from unauthorized access – highlighting that cybersecurity plays a practical protective role in today's world.

Cybercrime leads to the absence of peace

Peace is a necessity for human beings. It is the presence of safety, dignity and freedom – both online and offline. When hackers attack hospitals through data breaches, or when information leaks undetected, vulnerabilities set in and threaten democratic institutions, nations and overall human security.

In Southeast Asia alone, digital crimes are on the rise. These are not just your "ordinary" IT problems

– they are crimes that destabilize communities and affect the population through mass media communication. We need more than just technical fixes or support from local IT services, we need to promote cybersecurity in all sectors, including education.

Cybersecurity is a matter of justice

Justice in cyberspace means upholding digital rights: the right to privacy, access to information and protection from cybercrime.

> **“Victims of online abuse often experience delayed justice as perpetrators benefit from impunity due to limited cyber forensic capacity and resources.”**

Therefore, cybersecurity begins with all of us, and users can work together to strengthen cyber forensics through technological advancements.

As an aspiring law student, I believe justice systems must evolve to keep pace with constantly advancing technologies – not only to prosecute cybercrimes but also to prevent them through stronger frameworks, media and information literacy, cybersecurity education and public awareness. Cybersecurity is not just a technical field; it is an essential pillar of justice.

Bridging the gender gap in digital safety

The gender gap in cybersecurity is not about how women work in tech, but about who is most often targeted and labeled as vulnerable.

> **“Women and LGBTQ+ individuals face disproportionate online stalking, harassment and blackmail.”**

As a result, individuals may withdraw from digital platforms or face account suspension, which limits their access to opportunities.

However, solutions are emerging to better protect women and the LGBTQ+ individuals from digital harm; there are programs around the world that empower women with digital literacy and cyber defense skills. We must invest in these efforts globally, not only to close the gender gap from structural marginalization but to ensure a safe cyberspace and inclusive for all.

Youth engagement in cybersecurity

Youth are often seen as impulsive users of technology, but they are also among the most agile defenders. Across the globe, young people are launching online safety campaigns and helping spread awareness about how to maintain a safe and fair use of technology. We understand how platforms work because we grew up with them; our insight is invaluable.

Media and Information Literacy is an applied subject in the Senior High School curriculum under the K to 12 Basic Education Programme of the Department of Education in the Philippines. It is taught in both public and private institutions. This subject encouraged me, as a participative student, to assess and evaluate information whilst promoting online safety in the process. Digital literacy is an important skill that all users must acquire to develop critical thinking. I was motivated, as a youth advocate for digital literacy, to help younger people become responsible digital citizens. I believe that helping citizens and shaping them to become digitally literate brings cybersecurity and justice to light. I have seen school systems breached, and the administration struggle to retrieve learner and faculty information, as well as restore their reputation as institutions. Schools must secure their digital systems to maintain a safe and disciplined learning environment.

A call to action

Cybersecurity must be understood as a tool for peacebuilding and development, requiring a global commitment from all stakeholders, with the meaningful inclusion of women and youth.

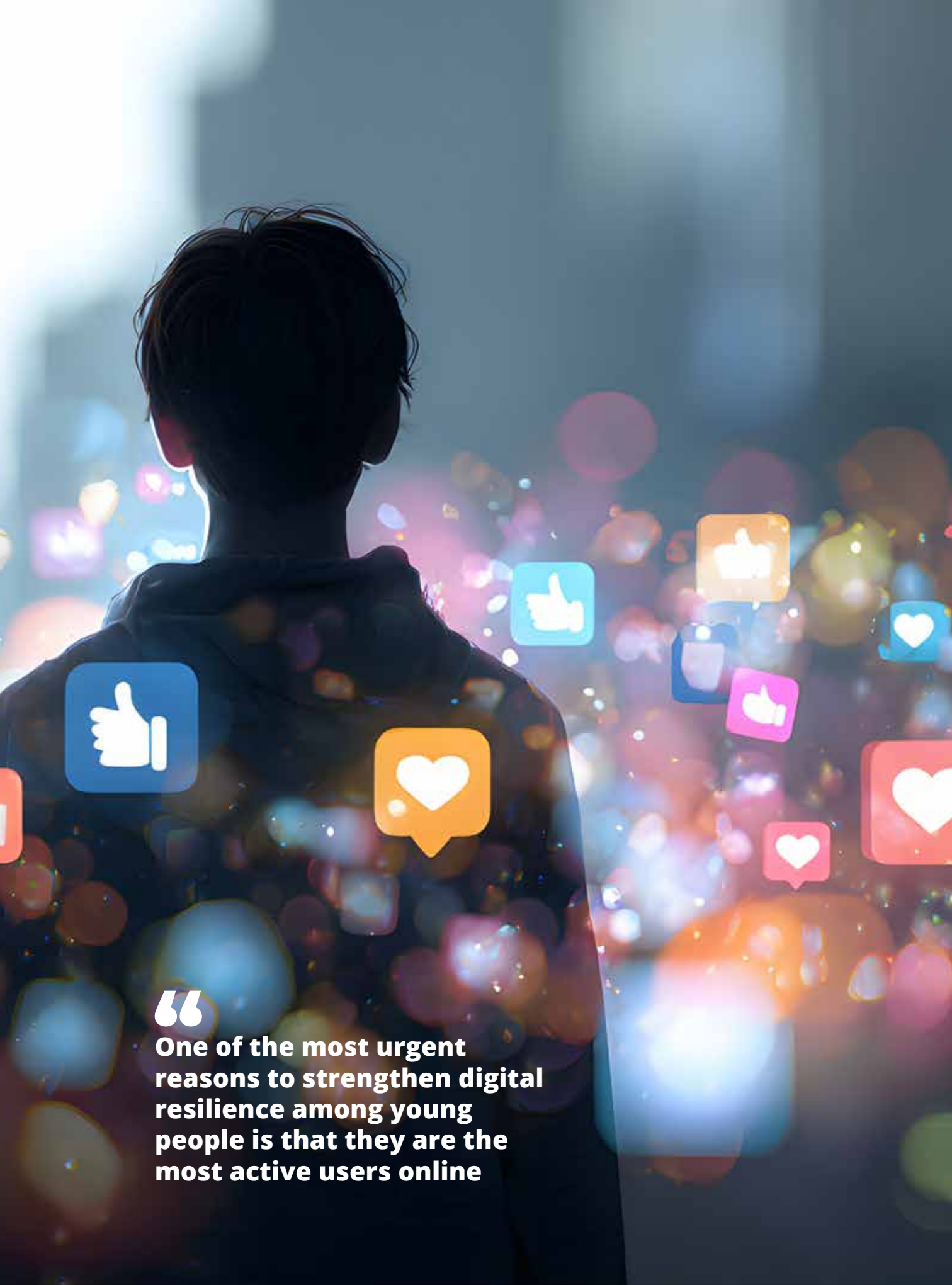
> **“We are in need of gender-inclusive digital safety policies, strong legal and educational frameworks against cybercrime and digital literacy programmes in schools and communities.”**

Cyber peace is not just an ideal to admire from afar; it is something we can advance through global cooperation.

Cyber peace begins with us. Our presence in this space is not optional. It is essential.

About the Author

Princess Tudtud is an aspiring law student and youth advocate from the Philippines, passionate about digital safety and youth empowerment. She has led school-based research and conducted interviews with professionals from SunStar Cebu, Cebu City Public Library and DAKILA – a Philippine Collective for Modern Heroism. Princess has also participated in local and educational programmes focusing on preserving cultural heritage and has contributed to promoting digital literacy and strengthening young people’s appreciation of history. Her writing reflects her vision and commitment to critical thinking.



One of the most urgent reasons to strengthen digital resilience among young people is that they are the most active users online

Strengthening digital resilience through youth empowerment: a global, youth-centered approach

by **Martina Matijević**

In today's interconnected world, the Internet has become the primary space for communication exchange, informal learning, entertainment, and identity formation – especially among youth. This growing digital immersion brings a range of risks, including cyberbullying, disinformation, online scams, exploitation, data misuse, and others. Yet digital resilience – the ability to navigate and recover from these risks – remains underdeveloped in many parts of the world.

One of the most urgent reasons to strengthen digital resilience among young people is that they are the most active users online. According to a study by Nagata et al. (2025), 69.5% of adolescents aged 11 to 15 have at least one social media account, including 63.8% of children under 13, despite minimum age restrictions. However, not all youth are equally protected or prepared to face cyber threats. In fact, a significant proportion of the world's youth live in developing countries where digital education often lags behind growing Internet access.

To address this imbalance and promote meaningful youth engagement, a comprehensive, multi-level strategy is required – one that goes beyond surface-level campaigns and empowers young people as agents of change. This article proposes two interlinked strategies: a three-phase, age-targeted institutional framework and the appointment of a UN Goodwill Ambassador for Digital Resilience.

A three-phase youth engagement framework

Digital resilience education must be adapted to the developmental stage of the learner. To do so, a three-phase model is proposed:

Phase 1: middle school students (ages 11-14)

In this phase, schools and police departments should collaborate to deliver interactive workshops

that teach students how to safely navigate social media, recognize potential online threats, and protect their privacy. Crucially, these workshops must be participatory rather than lecture-based. At the end of each workshop, students should be given a creative task (e.g., developing a poster campaign, conducting a peer survey, or creating a digital story) to complete over a set period.

Following this, police officers or digital safety educators would return to evaluate the students' progress and continue the discussion. Schools could also encourage students to organize public awareness events, allowing them to engage and educate their wider communities about digital resilience in an empowering and visible way.

Phase 2: high school students (ages 15–18)

At this stage, the content should evolve to reflect the growing complexity of adolescents' digital lives. Workshops should address more advanced topics such as cyberbullying, phishing, doxxing, and algorithmic manipulation. Students should take on leadership roles – for example, by forming digital safety clubs or leading peer-to-peer mentoring initiatives.

"Active involvement remains essential: learners must not simply absorb information, but apply it through debates, simulations, media content analysis (including misinformation and disinformation), and campaign development."

Phase 3: young adults (ages 19-24)

Universities, especially departments of computer science, psychology, education, and law, should play a proactive role in fostering digital resilience among youth. Courses dedicated to online ethics, privacy, digital rights, and cybersecurity should be integrated into undergraduate and non-degree programs. For students not enrolled in these courses, departments should offer regular workshops open to all disciplines.

Young adults should be encouraged to co-design these learning opportunities — whether by conducting research, facilitating discussions, or launching local digital literacy initiatives.

Appointing a UN Goodwill Ambassador for digital resilience

While grassroots and institutional efforts are essential, broader visibility can amplify their impact. A youth-centered, globally recognized public figure could play a valuable role, if selected carefully.

> **"The United Nations should consider appointing a Goodwill Ambassador for Digital Resilience, ideally someone from the youth demographic who has a strong and respected online presence."**

This ambassador should advocate for awareness of digital challenges and cyber threats affecting youth – such as online harassment, body shaming, or misinformation – through clear, authentic storytelling rather than technical jargon. Research shows that the mere-exposure effect – the tendency to be influenced by familiar figures – can make these messages more relatable, resonant, and memorable.

The chosen ambassador should engage their audience through informal, narrative-based education, using their platforms to raise awareness in an emotionally resonant way, rather than a commercially driven one.

This approach could ensure global reach and accessibility, especially when combined with localized programs in schools and universities. It also aligns with the goals of the 2030 Agenda for Sustainable Development, including quality education, reduced inequalities, and peaceful, inclusive societies.

Conclusion: a dual approach for a resilient future

The challenges of digital life cannot be addressed through awareness alone. They demand participation, structure, and representation. By combining institutional, age-sensitive education programs – especially in developing countries – with the outreach capacity of a well-selected Goodwill Ambassador, we can foster a global ecosystem of youth-led digital resilience.

This approach is not only about protecting youth from cyber threats; it is also about empowering them to recognize, resist, and reshape the digital world *together.*

About the Author

Martina Matijević is a law student from Croatia with a strong interest in the intersection of law, society, and human behaviour. She has authored several articles on human rights, corporate law, and the role of psychology in ensuring the effective and ethical functioning of legal practitioners. Her work reflects a commitment to exploring both the structural and human dimensions of legal systems, with the aim of promoting justice, accountability, and professional well-being.

United Nations
Interregional Crime and Justice
Research Institute

GHENT
UNIVERSITY