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, policy-makers, 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.