Risk management in social media: An educational framework for building digital security in the context of NIS2

Anita STOYANOVA[1], Emil DELINOV[2]

[1] Paisii Hilendarski University of Plovdiv,
Faculty of Mathematics and Informatics, Plovdiv, Bulgaria

[2] Trakia University, Department of Computer Science and Mathematics,
Stara Zagora, Bulgaria

anita.georgieva@gmail.com, emil.delinov@trakia-uni.bg

Abstract: *Social media platforms have become one of the central components of the contemporary digital environment, used for personal communication as well as for organizational and societal purposes. Alongside the opportunities they provide, these platforms also represent a significant source of information security risks, including social engineering, impersonation, disinformation, and data leakage. This paper examines risk management in social media in the context of Directive (EU) 2022/2555 (NIS2) and the guidelines of the European Union Agency for Cybersecurity (ENISA), considering social media as part of a broader digital ecosystem subject to systematic risk management. The main contribution of the study is educational and methodological. A model is proposed for integrating social media risk management into education, combining practice-oriented pedagogical approaches with a methodology for effectiveness measurement based on the HAIS-Q questionnaire*.

Keywords: Social media, Artificial Intelligence, Cybersecurity, Information security, NIS2, Risk management, Training, HAIS-Q.

1. Introduction

Over the past two decades, social media (SM) has transformed from a space for personal communication into a critical element of modern corporate and public digital infrastructure. SM platforms have become indispensable tools for business development, public communication, and crisis management. According to the classification by Kaplan & Haenlein (2010), these platforms—ranging from collaborative projects (Wikipedia) to social networking sites (Facebook, X, LinkedIn)—have become vital for stakeholder engagement. However, the mass integration of SM into the corporate and public sectors expands the attack surface by exploiting their informal nature and turning them into a serious vulnerability, leveraged through various tools such as phishing, disinformation, and identity theft (Boyd & Ellison, 2007).

https://doi.org/10.58503/icvl-v21y202613

1.1 Social media in the context of cybersecurity and NIS2

In response to evolving cyber threats, the European Union adopted Directive (EU) 2022/2555 (NIS2) in 2022 to harmonize cybersecurity across Member States (European Union, 2022). While NIS2 does not directly regulate social media, it treats them as part of an organization's digital ecosystem. Since these platforms are integrated into all business and management processes, they fall within the scope of mandatory risk management. Compromising institutional accounts can lead to reputational, financial, and legal consequences, threatening the overall resilience of the entities.

Risks in such an environment are hybrid—combining complex infrastructure with a powerful human factor (ENISA, 2023b). Blurred boundaries between personal and professional identity create challenges for control and compliance, turning SM into an attack surface where technical vulnerabilities intersect with psychological manipulation.

The unique attack surface presented by SM makes it both a critical digital vector and an interesting object for study and analysis. This paper examines the interaction between SM and information security, focusing on threats, providing examples of real incidents, and outlining educational applications in the context of NIS2.

2. Key threats and sectoral trends

This section presents the dominant threats, selected not through a formal taxonomy but based on the risk factor and the way SM amplifies their impact. The current analysis views dominant threats not as isolated problems but as a complex combination of behavioral and organizational factors.

2.1 Human factor related attacks

The human factor is the most frequently exploited vulnerability through psychological manipulation.

• **Social Engineering**: The most commonly exploited vulnerability, relying on trust and human habits rather than technical sophistication (Mitnick & Simon, 2002). Approximately 30% of social engineering campaigns target the general public, while 18% target public administration (ENISA, 2023b). They are most often associated with "external" phishing.

• **Internal Risks**: Threats arising from employee negligence (unintentional sharing of sensitive data, weak passwords) rather than targeted attacks. Research by Schultz et al. (2021) emphasizes that insiders remain one of the most persistent threats.

2.2 Information manipulation and disinformation

SM platforms enable the large-scale dissemination of false content to undermine trust—47% targeted at citizens and 29% at institutions, according to

data from (ENISA, 2023b). These threats blur the line between cybersecurity and information warfare.

An **example** is the Twitter breach of July 15, 2020, where spear-phishing was used to compromise over 130 official accounts (X, 2020; Isaac, Frenkel & Griffith, 2020).

2.3 AI-Supported attacks

Generative Artificial Intelligence (GenAI) facilitates the creation of convincing phishing messages and "deepfakes" (voice and video impersonation), which complicates the distinction between legitimate and malicious communication and necessitates a reassessment of trust mechanisms (ENISA, 2023b). This requires the adaptation of both technical safeguards and educational approaches.

2.4 Sectoral trends and context

For the period from July 2022 to June 2023, ENISA documented over 2,580 significant incidents. The most affected sector is public administration (19%), followed by healthcare (8%). In many of the aforementioned 2,500+ incidents, SM platforms acted as the initial contact channel or an amplifier of the impact.

The specificity of each sector requires adapted strategies, as universal measures are often insufficient against specific and cascading effects affecting multiple industrial, regional, national, and cross-border structures simultaneously.

3. Regulatory context: The NIS2 directive

The complexity of risks in social media necessitates the use of best practices and regulatory frameworks such as NIS2 for the structured management of digital risk. Although SM platforms are not directly regulated entities, their impact on operational resilience and the potential for data compromise means they must be addressed in every cybersecurity strategy (European Union, 2022). This section highlights some of the most relevant provisions of the Directive in this context.

• When SM are used for communication or public relations, they may fall under the regulation of **Article 21 (Risk Management)**, which obliges entities to implement “appropriate and proportionate technical, operational, and organizational measures.”

Organizations often use external tools to manage their social media presence—such as scheduling platforms, analytics dashboards, content generators, or influencer partnerships. NIS2 expands cybersecurity responsibility to include such service providers, treating them as part of the digital **Supply Chain (Article 22)**.

When activities in SM are linked to significant incidents affecting the availability of essential services or data confidentiality, they are subject to reporting to the national CSIRT (early warning within 24 hours) — **Incident Reporting (Articles 23 and 24)**.

According to **Article 21(2)**, training is key to mitigating the impact of the human factor. It is recommended to integrate practical scenarios into training curricula through simulated phishing via direct messages, analysis of real disinformation attacks, and tabletop exercises for identity theft response.

Such programs significantly impact the "human layer" of cybersecurity and support long-term organizational resilience (ENISA, 2021; ENISA, 2023c).

4. Methodological model for education in social media risk management and measurability (HAIS-Q)

Effective risk management in social media cannot be achieved solely through policies and technical controls. As demonstrated in the previous sections, a substantial proportion of threats originate from human behavior, decision-making processes, and limited contextual awareness when interacting with social platforms. Consequently, education and awareness-raising should be regarded as core components of sustainable cybersecurity, in line with the principles of the NIS2 Directive and the recommendations of ENISA.

This section proposes a methodological educational model aimed at students from different academic disciplines and educational levels. The model integrates theoretical knowledge, practical activities, and outcome measurement through the Human Aspects of Information Security Questionnaire (HAIS-Q).

4.1 Educational objectives and expected learning outcomes

The primary objective of the proposed model is to support students in developing sustainable knowledge, attitudes, and behavioral patterns related to social media risk management. The educational focus extends beyond threat recognition to fostering critical thinking, situational assessment, and the practical application of good security practices in real-world contexts.

The expected learning outcomes include:

- understanding the specific risks arising from the use of social media in organizational environments;
- the ability to identify scenarios involving social engineering, disinformation, and impersonation;
- applying fundamental risk management principles aligned with NIS2 requirements;
- developing responsible behavior in the use of social media in both personal and professional contexts.

4.2 Teaching and learning approaches

The proposed model relies on a combination of pedagogical approaches that reflect the dynamic and hybrid nature of social media environments. Core methods include:

- case-based learning, including the analysis of real-world incidents;
- scenario-based learning and tabletop exercises involving fake profiles, disinformation campaigns, and compromised accounts;
- discussions addressing ethical and reputational aspects of online behavior;
- hands-on tasks focused on configuring security settings (e.g., multi-factor authentication, access management);
- short critical-thinking assignments aimed at recognizing manipulative content.

This approach supports the transition from "knowing" to "doing" by placing students in an active role and encouraging reflection on the consequences of actions taken in social media environments.

4.3 Adaptation across disciplines and educational contexts

A key strength of the model is its multidisciplinary applicability. Social media risk topics can be integrated into courses such as information security, cybersecurity, social media and communications, internet-based technologies, public administration, as well as non-technical disciplines.

Depending on the academic profile, instructional emphasis may be placed on:

- risk analysis, policies, and regulatory compliance;
- threat modeling and social engineering techniques;
- reputational risk, ethics, and disinformation;
- digital identity, platform governance, and access control;
- media literacy and digital hygiene.

This flexibility allows the model to remain relevant across faculties and degree programs while preserving conceptual coherence.

4.4 Effectiveness assessment and continuous improvement using HAIS-Q

A central element of the proposed model is the use of HAIS-Q as a tool for evaluating educational effectiveness. HAIS-Q measures three core dimensions - knowledge, attitudes, and behavior - making it particularly suitable for educational environments.

Within the training process, the questionnaire may be used:

- as a pre-assessment tool to establish baseline awareness levels;
- as a post-assessment instrument to measure learning outcomes;
- to identify weaknesses across specific domains (e.g., phishing, password management, information sharing);
- as a basis for adapting and improving instructional content.

In this way, education becomes a measurable, iterative process of continuous improvement, focused on behavioral change rather than solely on theoretical knowledge acquisition.

4.5 Education as a factor of cyber resilience

Developing organizational resilience to social media-related threats is not purely a technical or regulatory challenge; it is fundamentally an educational one. The NIS2 Directive emphasizes the importance of training and awareness at all organizational levels, while ENISA analyses consistently identify human behavior as both a key vulnerability and the greatest opportunity for improvement.

In this context, education fulfills a dual role: it prepares future professionals to recognize and manage risks, and it fosters a culture of responsible and informed social media use. As such, education represents a long-term investment in the resilience and security of the digital environment.

5. Risk management approaches and good practices

The NIS2 Directive provides a regulatory framework, but its effective implementation depends on organizations adopting specific measures tailored to their context. Social media-related risks, due to their hybrid nature - technical, behavioral, and reputational - require an interdisciplinary approach that combines technical safeguards with policy, training, and culture-building.

The following use cases demonstrate how organizations across different sectors have approached social media risk in ways that reflect both the spirit of NIS2 and best practices recommended by ENISA and ISO/IEC 27001. These examples also serve as models for entities not formally covered by NIS2 but still highly exposed to social media threats.

5.1 Social media impersonation in the education sector

In a mid-sized secondary school, several teachers received what appeared to be an urgent request from the school principal via both Facebook Messenger and institutional email. The message asked for students' personal data under the pretense of updating emergency contact records. A vigilant staff member noticed subtle inconsistencies in the message's tone and phrasing, prompting an internal review.

It was discovered that a fraudulent Facebook profile had been created using the principal's public image and biographical information. The attacker combined this impersonation with spoofed internal emails - a hybrid attack exploiting both social media and traditional communication channels.

Although the school was not directly regulated under NIS2, it adopted several measures aligned with the directive's principles:

- Internal communication policy was revised to prohibit the exchange of sensitive data outside authenticated systems;
- All staff received targeted training in recognizing social engineering techniques;
- A verification protocol was established for unusual requests, including a mandatory second communication channel for confirmation.

This case reflects how non-regulated entities can voluntarily implement risk-aware behaviors in line with NIS2 Articles 21 and 24, particularly in the education sector, which ENISA classifies as a high-exposure environment (ENISA, 2023a).

5.2 Data exposure via linkedin in a professional services firm

A GDPR and ISO/IEC 27001-certified accounting firm experienced a reputational incident when an employee shared a celebratory post on LinkedIn. The accompanying photograph, taken at their workstation, inadvertently revealed a spreadsheet on a monitor containing sensitive client information (e.g., VAT numbers, internal identifiers).

Although the post was removed within hours, the incident highlighted a blind spot in the firm's digital hygiene. A root-cause analysis concluded that existing controls focused on internal systems but not on employee behavior in public digital spaces.

In response, the firm implemented a layered governance model for social media, including:

- A review and approval process for posts made from corporate devices or accounts;
- Mandatory training for all staff on oversharing, privacy by design, and personal branding ethics;
- Quarterly social media audits as part of the internal audit cycle.

This approach demonstrates the alignment of social media governance with Article 21 of NIS2 and Clause A.7 of ISO/IEC 27001, emphasizing awareness, access control, and acceptable use policies.

5.3 Social engineering against a small business through instagram spoofing

A small artisan enterprise, reliant on platforms like Instagram for client engagement and Stripe/PayPal for sales, received a message allegedly from Meta’s support team. The message claimed copyright infringement and urged the owner to log in via a provided link. Although the login page was a spoof, the business was protected by two-factor authentication (2FA), which prevented account takeover.

The event served as a wake-up call for the owner, who took the following steps:

- Created an asset inventory including social media accounts;
- Implemented password management using a secure vault;
- Established weekly backup procedures for web and social media content;
- Drafted an incident checklist for social engineering attempts.

This case reflects the relevance of ENISA’s Cybersecurity Guide for SMEs (ENISA, 2022b), and demonstrates how NIS2 principles (risk identification, response preparedness, and cultural awareness) can be adapted at a micro-enterprise scale.

5.4 Organizational social media checklist

Beyond formal regulation and sectoral trends, organizations need actionable tools to assess their internal readiness for secure social media use. A checklist is an accessible, adaptable mechanism that supports continuous improvement and operational alignment with both NIS2 expectations and general cybersecurity frameworks such as ISO/IEC 27001.

Such tools are particularly useful for SMEs and non-regulated entities, offering a structured way to identify vulnerabilities, enforce accountability, and prioritize remediation. This approach reflects the risk-based methodology promoted in Article 21 of the NIS2 Directive and ENISA's guidance on establishing cybersecurity maturity (ENISA, 2022a), (ENISA, 2022b).

5.5 Social media governance self-assessment questions:

- Do we have a documented and up-to-date social media policy?
- Are roles and responsibilities clearly defined for account access and content management?
- Do all organizational accounts use strong passwords and multi-factor authentication (MFA)?
- Are posts from official channels reviewed or approved before publication?
- Is there a process for monitoring account activity and identifying anomalies?
- Are employees trained regularly on social media risks such as phishing and oversharing?
- Is social media included in the organization's incident response plan?
- Have we reviewed the security posture of third-party platforms (e.g., scheduling, analytics)?
- Do we have procedures for revoking access when employees leave or roles change?
- Is there a designated point of contact responsible for coordinating digital communications and security?

This checklist can serve as the foundation for more advanced governance tools, such as internal audits, KPIs for cybersecurity posture, or integration into broader digital risk assessments. When used proactively, it transforms social media from a liability into a managed and measurable component of the organization's digital ecosystem.

6. Teaching and awareness implications

Developing organizational resilience to social media threats is not solely a technical or regulatory matter - it is fundamentally educational. NIS2 emphasizes the importance of training and awareness across all organizational levels (Article 21), while ENISA consistently identifies human behavior as both a key

vulnerability and the greatest opportunity for improvement (ENISA, 2022a), (ENISA, 2023c).

Social media poses unique pedagogical challenges and opportunities. Its informal tone, dynamic pace, and blending of personal and professional boundaries make it difficult to govern using static policies alone. Education must therefore equip individuals - not only security professionals, but also communication officers, HR staff, and students - with the skills to recognize, assess, and respond to risks in these environments.

6.1 Integrating social media security into cybersecurity curricula

Formal cybersecurity training and academic programs increasingly include modules on risk governance, digital identity, and online manipulation. However, social media is often treated as a peripheral topic. Given its growing relevance, educators should embed it directly into core curricula, using real-world incidents and simulations to drive engagement and critical reflection.

For example, case studies like the 2020 Twitter breach or Instagram phishing messages can be used to:

- Analyze how technical and human failures interact;
- Practice incident response planning and root cause analysis;
- Discuss organizational accountability and reputational risk.

ENISA's Cybersecurity Training Guidelines (ENISA, 2021) recommend scenario-based learning, cross-functional exercises, and practical tool use - strategies well-suited to teaching social media risk management.

6.2 Targeting learners at multiple levels

Cybersecurity awareness must be tailored to the audience. For younger students, especially in secondary and early tertiary education, media literacy and personal digital hygiene are foundational. OECD studies emphasize that empowering young users to critically evaluate online content is essential to long-term security culture (OECD, 2021).

For university students and professionals, especially in IT, law, journalism, or public administration, more advanced competencies are needed. These may include:

- Threat modeling for public-facing accounts;
- Regulatory compliance for content and data;
- Ethical guidelines for institutional communications.

The European Commission's DigComp framework (European Commission, 2019) outlines five dimensions of digital competence - information literacy, communication, content creation, safety, and problem-solving - all of which are relevant when navigating social media securely.

6.3 Measuring training effectiveness with hais-q

The HAIS-Q model includes seven behavioral domains (e.g., password management, internet use, social networking) and evaluates three dimensions for

each: knowledge, attitude, and behavior. Participants respond using a five-point Likert scale ranging from "strongly disagree" to "strongly agree."

This tool allows organizations to benchmark their staff's digital risk awareness before and after training and identify domain-specific weaknesses (e.g., oversharing on LinkedIn);

The final purpose is not only to gauge awareness but also to shape behavior over time through reflection, feedback, and policy alignment.

6.4 Creating a culture of shared responsibility

Organizations that foster a security culture - where every employee understands their role in protecting digital assets - report significantly fewer incidents caused by negligence (ENISA, 2023c), (Schultz, Proctor, Lien & Salvendy, 2021). Social media presents a clear opportunity to reinforce this culture, as nearly every department interacts with these platforms.

To that end, educational efforts should include:

- Policy co-creation workshops involving staff across departments;
- Regular phishing simulations using social messaging contexts;
- Story-based learning, using internal or public incidents to foster discussion.

Ultimately, the goal is to shift from compliance-driven behavior to proactive digital citizenship, where individuals feel both empowered and accountable. As ENISA's 2023 culture guidelines state, "Awareness is not the goal - behavioral change is" (ENISA, 2023c).

7. Conclusion

Social media has become a defining element of modern organizational life, serving simultaneously as a critical communication channel and an expanding vector for cyber threats. As demonstrated throughout this paper, the integration of social platforms into business, education, public administration, and civic engagement creates a dual reality in which opportunities for outreach, visibility, and dialogue coexist with vulnerabilities related to data protection, impersonation, disinformation, and human error.

The evolving threat landscape, as documented by ENISA and related studies, increasingly exploits openness, immediacy, and trust-based interactions characteristic of social media environments. In such conditions, traditional technical controls alone are insufficient unless complemented by coherent policies, continuous training, and risk awareness across all organizational roles.

Although the NIS2 Directive does not explicitly regulate social media platforms, it establishes legal and organizational obligations that require entities to address their digital presence in a holistic manner. Incident response preparedness, staff training, and third-party risk management all implicitly encompass social media governance. The case studies and sectoral practices presented in this paper demonstrate that effective social media risk management is achievable across

diverse organizational contexts and sizes, if strategies are aligned with a culture of cybersecurity and shared responsibility.

Education emerges as a central enabler of this transition. By embedding social media risk topics into cybersecurity and related curricula, promoting awareness at multiple organizational levels, and modeling responsible digital behavior, both educational institutions and organizations can strengthen long-term cyber resilience.

Ultimately, the ability to leverage the benefits of social media while systematically managing its risks will become a defining capability of digitally mature and resilient institutions.

REFERENCES

Boyd, D., & Ellison, N. (2007) Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*. 13(1), 210-230.

ENISA. (2021) *Cybersecurity Training - Guidelines for Education Providers*. https://www.enisa.europa.eu/publications/cybersecurity-training-guidelines-for-education-providers [Accessed: 15 July 2022].

ENISA. (2022a) *Guidelines on Cybersecurity Culture for Organisations*. https://www.enisa.europa.eu/publications/guidelines-on-cybersecurity-culture-for-organisations [Accessed: 18 June 2023].

ENISA. (2022b) *Cybersecurity Skills Development in the EU*. https://www.enisa.europa.eu/publications/cybersecurity-skills-development-in-the-eu [Accessed: 21 Jan 2023].

ENISA. (2022c) *ENISA Threat Landscape for Education Sector*. https://www.enisa.europa.eu/publications/enisa-threat-landscape-for-education-sector [Accessed: 18 May 2023].

ENISA. (2023a) *ENISA Threat Landscape 2023*. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023 [Accessed: 7 May 2024].

ENISA. (2023b) *Cybersecurity Culture Guidelines: Behavioural Aspects of Cybersecurity.* https://www.enisa.europa.eu/publications/cybersecurity-culture-guidelines [Accessed: 15 July 2024].

European Commission. (2019) *The Digital Competence Framework for Citizens (DigComp).* https://joint-research-centre.ec.europa.eu/digcomp_en [Accessed: 18 July 2025].

European Union. (2022) *Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS2 Directive).* https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng [Accessed: 18 July 2025].

Isaac, M., Frenkel, S., & Griffith, E. (2020) Twitter Hack Exposes Broader Risk to Internet Security. *The New York Times*. July 16. https://www.nytimes.com/2020/07/16/technology/twitter-hack.html.

Kaplan, A. M. & Haenlein, M. (2010) Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*. 53(1), 59-68.

Mitnick, K. D. & Simon, W. L. (2002) *The Art of Deception: Controlling the Human Element of Security*. Wiley.

OECD. (2021) *Empowering Young Children in the Digital Age*. https://www.oecd.org/education/empowering-young-children-in-the-digital-age.htm [Accessed: 18 July 2025].

Parsons et al. (2017) The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies. *Computers & Security*. 66, 40-51. https://doi.org/10.1016/j.cose.2017.01.004.

Schultz, E. E., Proctor, R. W., Lien, M. C. & Salvendy, G. (2021) Human factors in information security: The insider threat-who can you trust these days? *Ergonomics in Design*. 29(1), 4-11.

Wardle, C. & Derakhshan, H. (2017) Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making. *Council of Europe report DGI* (2017).

X. (2020) An update on our security incident. https://blog.x.com/en_us/topics/company/2020/an-update-on-our-security-incident [Accessed: 18 July 2025].