Human Factors in Cybersecurity, Vol. 208, 2026, 106–116
https://doi.org/10.54941/ahfe1007417

Designing an AI-Driven Framework for Human-Centered Cybersecurity Practices

Kitty Kioskli[1], Pedro Tomás[2], Wissam Mallouli[3], João Fernandes[2], Dimitrios Koutras[4], Luis Cordeiro[2], and Dimitrios Kallergis[5]

[1]Trustilio B. V., Vijzelstraat 68, 1017 HL Amsterdam, The Netherlands
[2]OneSource Consultoria Informática Lda, Rua Dom João de Castro, Coimbra, Portugal
[3]Montimage, 39 Rue Bobillot, 75013 Paris, France
[4]Department of Informatics, University of Piraeus, 185 34 Piraeus, Greece
[5]Department of Informatics and Computer Engineering, University of West Attica, 122 43, Aegaleo, Greece

ABSTRACT

Cybersecurity systems are often fragmented and difficult to navigate, leaving organisations particularly small and medium-sized enterprises (SMEs) struggling to implement effective, human-centered, and resilient security practices. End users face dispersed resources, complex regulatory requirements, and limited practical guidance, resulting in uneven levels of preparedness and cyber hygiene. These gaps undermine decision-making, organisational resilience, and the effectiveness of certification and compliance processes. To address these challenges, the paper proposes a holistic, conceptual framework that integrates human-centered principles with explainable artificial intelligence (AI) and system-level collaboration. Drawing on established approaches in human-centered security, privacy-by-design, resilience engineering, regulatory science, and AI-driven decision support, the framework aligns with the European Cybersecurity Skills Framework (ECSF). It synthesises insights from cross-sector analyses, socio-technical modelling, and European cybersecurity initiatives that emphasise interoperability and human factors. The framework is structured around five interconnected components: (i) a human-centered decision-support layer using explainable AI; (ii) a harmonised catalogue of cybersecurity, training, and regulatory resources; (iii) an interoperability and collaboration layer enabling structured, machine-readable information exchange; (iv) an adaptive learning and training component aligned with behavioural and competency models; and (v) a trust-by-design compliance engine supporting certification and conformity assessment. The analysis shows that combining human factors with explainable AI produces clearer, more actionable guidance with the potential to reduce cognitive and operational burden, subject to user-centered design and validation. Interoperability and collaboration mechanisms help overcome fragmentation, while adaptive learning pathways tailor support to skill levels and organisational maturity. Overall, the framework reframes cybersecurity as a socio-technical system shaped by people, regulation, and collaboration. Future work will empirically validate the framework across diverse organisational contexts to assess its practical impact.

Keywords: Human-centered cybersecurity, Explainable AI, Socio-technical systems, Cyber resilience, Certification processes, Behavioural cybersecurity, Adaptive training, Collaborative frameworks

Received January 4, 2026; Revised April 5, 2026; Accepted April 25, 2026; Available online July 20, 2026

INTRODUCTION

Rapid digitalisation has produced increasingly complex cybersecurity ecosystems, where vulnerabilities stem from technical flaws, organisational practices, human factors, and fragmented governance. In healthcare, for instance, heterogeneous stakeholders, legacy infrastructures, and uneven cyber hygiene practices create systemic exposure to cyber threats, with cascading impacts on privacy, safety, and service continuity (Kioskli et al., 2022; Kioskli et al., 2023). At the same time, human factors research confirms that user attitudes, risk perceptions, and decision-making strategies are central determinants of cybersecurity outcomes (Almansoori et al., 2024).

The European regulatory landscape has evolved rapidly, emphasising resilience, risk management, and cross-sector assurance. The NIS2 Directive strengthens obligations for essential and important entities, encompassing risk management, incident reporting, and governance responsibilities (Directive (EU) 2022/2555). The Cyber Resilience Act introduces horizontal cybersecurity requirements for products with digital elements, targeting persistent weaknesses in product security and lifecycle vulnerability management, while the AI Act establishes harmonised rules for trustworthy AI — covering safety, transparency, and human oversight (Regulation (EU) 2024/2847; Regulation (EU) 2024/1689). To address persistent skills and capacity gaps, the European Cybersecurity Skills Framework provides a common taxonomy for cybersecurity roles, competences, and training, fostering coherent workforce development across Member States (ENISA, 2022). Against this backdrop, the need for human-centered, socio-technical approaches integrating explainable AI (XAI), behavioural insights, and regulatory alignment has become increasingly apparent (Hakami & Alshaikh, 2022; Almansoori et al., 2023).

In this context, we aim to design a conceptual AI-driven framework for human-centered cybersecurity practices that can inform the future development of decision-support tools and collaborative infrastructures. Building on prior work on human-centric cyber hygiene and security-by-design in complex ecosystems (Kioskli et al., 2022, 2023), and on advances in explainable AI and human-centered AI design (e.g., Samek et al., 2019; Shneiderman, 2022), we pursue three main objectives. First, we articulate a methodological approach for deriving framework requirements by combining socio-technical analysis, regulatory drivers, and human-factor considerations. Second, we propose a multi-layered AI-enabled framework that integrates human-centered decision-support, shared knowledge resources, interoperability and collaboration mechanisms, adaptive learning and behavioural support, and compliance-by-design elements. Third, we offer a conceptual analysis of how such a framework could reduce fragmentation, support more consistent cyber hygiene and facilitate alignment with evolving regulatory and certification processes across sectors. The overall goal is to provide a theoretically grounded, human-centered blueprint that can guide future empirical validation and implementation studies in real-world cybersecurity ecosystems.

The framework is intentionally designed as a high-level, integrative model, with the expectation that individual components will be further specialised and implemented in domain-specific contexts.

Methodological Approach for Framework Design

The methodological approach for designing the proposed framework follows a structured socio-technical process. The initial step involved is to analyse the persistent dispersion of cybersecurity tools, training resources, and regulatory documentation across disconnected repositories. This dispersion, combined with inconsistent taxonomies and limited semantic descriptions, restricts the ability of organisations particularly small- and medium-sized enterprises (SMEs), security operations centers (SOCs), and computer security incident response teams (CSIRTs) to effectively identify, compare, and operationalise cybersecurity resources. Insights from recent systematic analyses of cybersecurity readiness and organisational maturity underscore that such fragmentation directly undermines the adoption of secure practices and reduces the efficiency of cyber risk management (Huarng et al., 2023). These findings informed the identification of core methodological requirements: resource harmonisation, human-centeredness, explainability, and interoperability.

Building on this needs analysis, the methodological design incorporated socio-technical cybersecurity principles to ensure that the framework adequately supports real-world decision-making. Empirical studies highlight that cybersecurity behaviour is shaped not only by technical controls but also by cognitive load, organisational context, and intentions (Alshaikh, 2020; Weickert et al., 2023). These insights were used to inform the design of human-centered reasoning components and adaptive behavioural support mechanisms within the framework. Complementing this perspective, contemporary work on cybersecurity governance emphasises the role of clear communication structures, well-defined responsibilities, and collaborative support environments in achieving meaningful operational security improvements (Stoleriu et al., 2025). These findings guided the inclusion of collaboration-oriented design requirements, ensuring that the framework facilitates cross-organisational alignment and shared situational awareness.

A further methodological pillar involved is to evaluate state-of-the-art developments in XAI for determining how transparent and trustworthy AI methods can support security-relevant decisions. Recent systematic reviews illustrate that explainable AI enhances interpretability, supports regulatory alignment, and increases user trust when deployed in sensitive socio-technical environments (Carvalho et al., 2019; Vilone & Longo, 2021). These insights informed the framework's emphasis on explainable reasoning, the integration of auditable decision pathways, and the alignment of AI outputs with user needs and regulatory requirements. The methodology also drew upon ENISA's recent guidance on securing AI systems, which stresses the need for transparency, human oversight, and risk-aware deployment practices in cybersecurity contexts (ENISA, 2023). Together, these considerations supported the development of a reasoning layer capable of delivering context-aware, interpretable recommendations.

Finally, the methodological approach emphasised interoperability and structured information exchange as essential elements for supporting real-world adoption. Persistent communication silos across European cybersecurity stakeholders have been shown to hinder coordinated threat

response and reduce the utility of shared intelligence (van Haastrecht et al., 2021). To address this, the methodology incorporated standards-aligned data modelling principles and machine-readable interaction mechanisms inspired by ENISA's interoperability work (ENISA, 2023). By synthesising these socio-technical, regulatory, and AI-driven insights, the methodology established a coherent foundation for the proposed AI-enabled, human-centered cybersecurity framework.

The AI-Driven Human-Centered Cybersecurity Framework

The framework is a multi-layered socio-technical ecosystem consolidating cybersecurity tools, training, legislation, and standards into a unified, human-centered environment—integrating discovery, decision-support, and interoperability to overcome fragmentation across technical, organisational, and human dimensions. As illustrated in Figure 1, the system is structured around a modular design that enables personalised interaction, secure information exchange, and continuous knowledge evolution. Table 1 presents further details about the framework core components.

Table 1: Central components and functions of the proposed cybersecurity framework.

Framework Component	Core Functions	Key Inputs	Outputs / Contribution
Human-Centered AI Decision Support	Interprets user intent; provides contextualised, explainable guidance	User queries; organisational profile; sector context; user previous interactions	Actionable recommendations; reduced cognitive burden
Harmonised Knowledge Catalogue	Aggregates tools, training resources, legislation, and standards	Validated resources; metadata; taxonomies	Unified search; semantic alignment; improved discoverability
Interoperability & Collaboration Layer	Enables secure, structured exchanges across entities and systems	External tools/services; secure communication protocols	Trusted data flows; coordinated incident response
Adaptive Learning & Support	Delivers personalised training pathways; integrates behavioural insights	Skill frameworks; user performance; behavioural indicators	Enhanced cyber hygiene; capacity building
Compliance-by-Design Engine	Aligns organisational controls with regulatory and certification requirements	Legislative frameworks; standardisation models	Automated compliance checks; traceability; readiness support

The framework centres on an explainable AI decision-support environment that interprets user needs — contextualised by sector, maturity, and compliance obligations — and guides identification of appropriate tools, training, and regulatory requirements through a unified, semantically

structured knowledge catalogue. A dedicated interoperability layer enables secure, standard-aligned communication with external stakeholders, while adaptive learning and compliance-by-design mechanisms deliver personalised training pathways and automated regulatory alignment checks, supporting informed decision-making and continuous preparedness within a federated, human-centered cybersecurity ecosystem.

Figure 1 illustrates the modular architecture of the proposed AI-driven, human-centered cybersecurity framework, highlighting the layered interaction between decision-support, harmonised knowledge resources, interoperability mechanisms, adaptive learning functions, and compliance-by-design components.

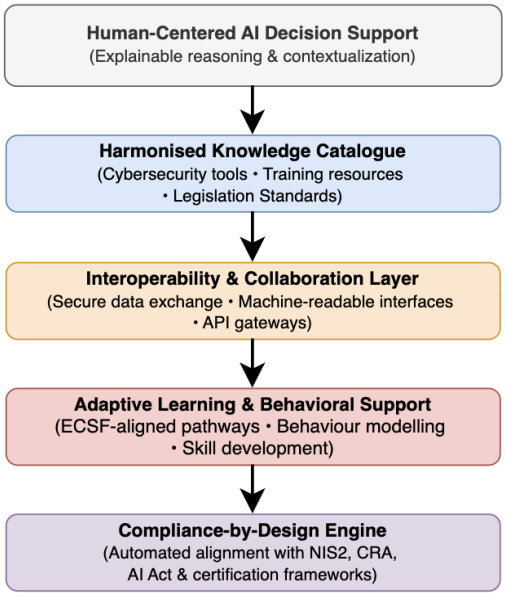

Figure 1: High-level architecture of the proposed cybersecurity framework.

Operationalisation and System Workflow

The proposed framework follows a structured operational workflow. Users submit a natural language or structured query, which the AI decision-support layer processes through intent recognition, contextual enrichment, and query decomposition.

The enriched query is matched against a harmonised knowledge catalogue, where relevant resources are semantically retrieved and ranked. An explainable reasoning engine then generates transparent recommendations, complete with justification paths and confidence indicators.

An interoperability layer enables optional integration with external systems — such as SOC platforms, compliance tools, or threat intelligence feeds — for real-time data enrichment. Outputs are subsequently passed to adaptive learning and compliance modules, which recommend targeted training interventions and perform alignment checks against machine-readable regulatory requirements.

User feedback and interaction logs are continuously reintegrated into the system, enabling iterative refinement of recommendations, personalisation models, and knowledge resources. This closed-loop design ensures the framework evolves alongside user needs and regulatory developments, and supports future implementation as a modular, service-oriented architecture.

Figure 2 illustrates the end-to-end operational workflow of the proposed framework, highlighting how user input is transformed into context-aware, explainable cybersecurity guidance and providing a step-by-step representation of components interaction and data flow across the system. The process begins with intent analysis and contextual enrichment within the human-centered AI decision-support layer, followed by semantic retrieval from the harmonised knowledge catalogue. The interoperability layer enables integration with external systems and data sources, while the adaptive learning and compliance components translate recommendations into personalised training actions and regulatory alignment checks. Continuous feedback loops support iterative refinement of system outputs and user interaction models, ensuring that the framework remains responsive to evolving organisational needs and cybersecurity contexts.

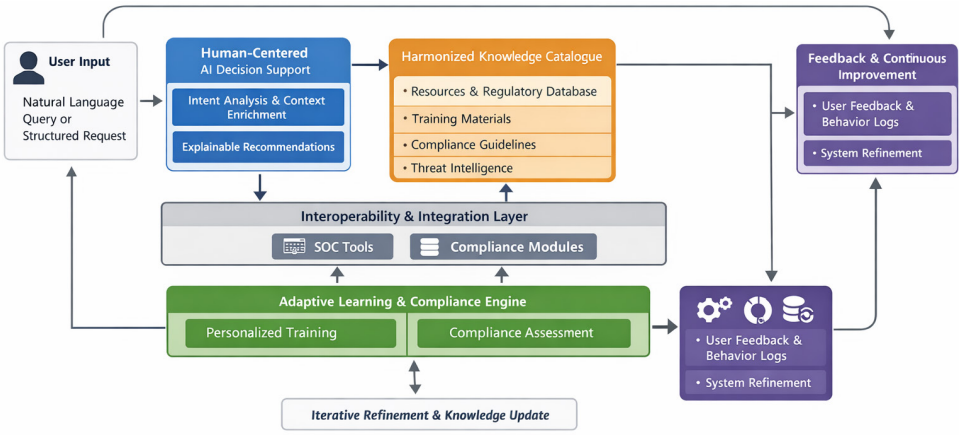

Figure 2: Operational workflow of the proposed AI-driven human-centered cybersecurity framework.

Human-Centered AI Decision & Support and Knowledge Resources

The human-centered AI decision-support layer provides transparent, contextualised guidance for navigating complex cybersecurity tasks. By interpreting natural language and structured queries, the system tailors the recommendations to sectoral characteristics, organisational maturity, and regulatory obligations — reducing cognitive burden and supporting informed decision-making across varying levels of expertise. Human-in-the-loop mechanisms and interpretable reasoning pathways ensure traceability and consistency across diverse operational environments.

Underpinning this capability is a harmonised knowledge repository consolidating cybersecurity tools, training resources, regulatory frameworks, and standards into a unified semantic structure. Dynamic interaction between both components creates an iterative refinement cycle, ensuring

recommendations evolve alongside technological advances, regulatory updates, and emerging threats — strengthening organisational preparedness and cybersecurity competency development.

Interoperability, Collaboration, and Secure Information Exchange

Interoperability and secure collaboration constitute central elements of the proposed framework, addressing long-standing challenges associated with fragmented communication channels and limited cross-organisational coordination. The interoperability layer addresses fragmented communication and limited cross-organisational coordination by enabling structured, machine-readable interaction between internal components and external stakeholders — including SOCs, CSIRTs, SMEs, and regulatory bodies. Standard-aligned communication models support capability negotiation, contextual information sharing, and seamless integration of external tools. A comprehensive authentication, authorisation, and accounting mechanism enforces secure access control and maintains auditability across all interactions, ensuring compliance with sector-specific governance requirements.

This collaborative environment facilitates timely detection, informed decision-making, and coordinated incident response. By accommodating heterogeneous systems and legacy protocols, the framework reduces operational fragmentation, fosters collective resilience, and supports scalable, real-time cybersecurity cooperation aligned with evolving regulatory landscapes.

Adaptive Learning, Behavioural Support, and Compliance-by-Design

This component integrates adaptive learning mechanisms with behavioural support functions to strengthen cybersecurity skills and organisational readiness. Personalised training pathways align with established competency structures and respond dynamically to user performance and sectoral maturity. By embedding feedback loops, the system reduces cognitive burden, promotes cyber hygiene, and ensures users can effectively operationalise secure practices.

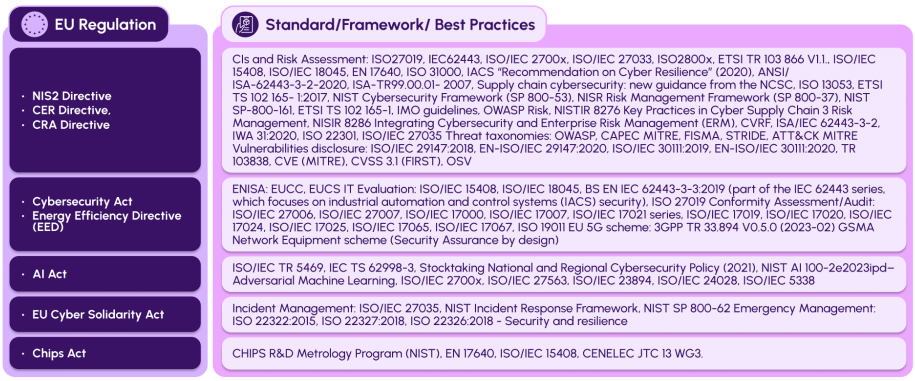

Figure 3: Regulatory and standardisation landscape.

A compliance-by-design engine embeds regulatory and standardisation requirements (Figure 3) directly into the reasoning process, structuring legislative obligations and certification criteria into machine-readable representations for automated alignment checks. Combined with adaptive learning, this ensures recommended actions, tools, and training outputs remain consistent with evolving governance requirements — supporting both continuous capacity building and sustained regulatory conformity.

Conceptual Analysis and Expected Outcomes

Integrating human-centered AI reasoning with a harmonised knowledge base reduces cognitive burden for cybersecurity stakeholders, delivering contextualised, transparent guidance that addresses usability challenges and maps user needs to relevant tools, training resources, and regulatory obligations.

Interoperability functions enable secure, auditable information exchange across organisational boundaries, improving situational awareness and supporting coordinated threat response, while integration of heterogeneous tools mitigates fragmentation and facilitates alignment of existing cybersecurity assets.

Finally, adaptive learning and compliance-by-design mechanisms strengthen organisational readiness and long-term resilience, promoting continuous capacity building, regulatory alignment, and a scalable foundation for human-centered cybersecurity resilience across diverse operational contexts.

These expected benefits remain to be empirically validated and depend on the appropriate design of explanation mechanisms and user interaction models.

Limitations and Trade-offs

While the proposed framework provides a comprehensive and human-centered approach to cybersecurity support, several limitations and trade-offs must be acknowledged.

First, the framework is conceptual and has not yet been empirically validated in real-world environments. As such, claims regarding reduced cognitive burden, improved decision-making, and enhanced resilience should be interpreted as theoretically grounded but not yet experimentally confirmed.

Second, the use of explainable AI introduces a potential trade-off between transparency and cognitive load. While explainability can enhance trust and understanding, excessive or poorly designed explanations may overwhelm users, particularly those with limited expertise. To mitigate this risk, the framework assumes the use of adaptive explanation mechanisms that tailor the level of detail to user profiles and contextual needs.

Third, the integration of heterogeneous data sources and interoperability mechanisms may introduce technical and organisational complexity, including challenges related to data standardisation, governance, and trust management across entities.

Finally, the broad scope of the framework—spanning decision support, training, interoperability, and compliance—may limit the depth of implementation in early-stage deployments. Future work should prioritise incremental implementation and domain-specific validation to ensure practical feasibility.

These considerations highlight the need for careful, context-aware deployment and systematic evaluation in future research.

CONCLUSION

The proposed framework, aligned with the European Cybersecurity Skills Framework (ECSF), presents a structured and human-centered response to fragmentation, usability barriers, and compliance complexity in modern cybersecurity ecosystems. By combining explainable AI-based decision support, a harmonised knowledge repository, interoperable communication mechanisms, adaptive learning pathways, and compliance-by-design principles, it offers a coherent model for improving organisational preparedness and resilience. The framework embeds socio-technical and behavioural considerations through transparency, contextualisation, and continuous knowledge evolution, enabling clearer decision-making, more effective resource use, reduced operational burden, and more consistent cyber hygiene across organisations. While conceptually robust, the framework remains theoretical and requires empirical validation across diverse sectors and contexts.

To support future empirical validation, a structured evaluation strategy is envisaged. This includes controlled user studies assessing usability, cognitive load, and trust in explainable recommendations, as well as organisational case studies measuring improvements in cyber hygiene, decision accuracy, and response times. Quantitative metrics such as task completion time, error rates, and compliance alignment scores will be combined with qualitative feedback to assess user experience and perceived value. In addition, pilot deployments in SME and sector-specific environments will be used to evaluate interoperability performance, integration complexity, and scalability of the framework. Comparative analyses with existing cybersecurity support tools will further help quantify its added value. This multi-method evaluation approach will provide the empirical evidence necessary to validate the framework's effectiveness and to refine its design for real-world adoption.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support provided for the following projects: the 'Advanced Cybersecurity Awareness Ecosystem for SMEs' (NERO) project, which has received funding from the European Union's DEP programme under grant agreement No. 101127411 and 'the 'Harmonizing People, Processes, and Technology for Robust Cybersecurity' (CyberSynchrony) project, which has received funding by the European Union's Digital Europe programme under Grant Agreement no. 101158555 and supported by the European Cybersecurity Competence Centre (ECCC). The

views expressed in this paper represent only the views of the authors and not those of the European Commission or the partners in the above-mentioned projects. Finally, the authors declare that there are no conflicts of interest, including any financial or personal relationships, that could be perceived as potential conflicts.

REFERENCES

Almansoori, A., Al-Emran, M., and Shaalan, K. (2023). Exploring the frontiers of cybersecurity behavior: A systematic review of studies and theories. Applied Sciences, 13(9), 5700. https://doi.org/10.3390/app13095700

Alshaikh, M. (2020). Developing cybersecurity culture to influence employee behavior: A practice perspective. Computers & Security, 98, 102003. https://doi.org/10.1016/j.cose.2020.102003

Baltuttis, D., Teubner, T., and Adam, M.T.P. (2024). A typology of cybersecurity behavior among knowledge workers. Computers & Security, 140, 103741. https://doi.org/10.1016/j.cose.2024.103741

Carvalho, D.V., Pereira, E.M., and Cardoso, J.S. (2019). Machine learning interpretability: A survey on methods and metrics. Electronics, 8(8), 832. https://doi.org/10.3390/electronics8080832

Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive) (Text with EEA relevance). Official Journal of the European Union, L 333, 27.12.2022, pp. 80–152.

ENISA. (2022). European Cybersecurity Skills Framework (ECSF) – User manual. European Union Agency for Cybersecurity.

ENISA. (2023). Multilayer Framework for Good Cybersecurity Practices for AI. European Union Agency for Cybersecurity (ENISA). Retrieved from https://www.enisa.europa.eu/publications/multilayer-framework-for-good-cybersecurity-practices-for-ai

European Union Agency for Cybersecurity (ENISA). (2022). European Cybersecurity Skills Framework (ECSF): User manual. ENISA. https://doi.org/10.2824/859537

European Parliament and Council. (2024a). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Official Journal of the European Union.

European Parliament and Council. (2024b). Regulation (EU) 2024/2847 on horizontal cybersecurity requirements for products with digital elements and amending Regulations (EU) No 168/2013 and (EU) 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act) . Official Journal of the European Union.

Hakami, M. and Alshaikh, M. (2022). Identifying strategies to address human cybersecurity behavior: A review study. International Journal of Computer Science and Network Security, 22(4), 299–309. https://doi.org/10.22937/ijcsns.2022.22.4.37

Huarng, K.-H., Lee, C.-F., & Yu, T. H.-K. (2023). Case study of a healthcare virtual community model. Technological Forecasting and Social Change, 188, Article 122281. https://doi.org/10.1016/j.techfore.2022.122281

Kioskli, K., Dellagiacoma, D., Fotis, T., and Mouratidis, H. (2022). The supply chain of a Living Lab: Modelling security, privacy, and vulnerability issues and potential mitigation strategies. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 13(2), 147–182. https://doi.org/10.22667/JOWUA.2022.06.30.147

Kioskli, K., Fotis, T., Nifakos, S., and Mouratidis, H. (2023). The importance of conceptualising the human-centric approach in maintaining and promoting cybersecurity hygiene in Healthcare 4.0. Applied Sciences, 13(6), 3410. https://doi.org/10.3390/app13063410

van Haastrecht, M., Golpur, G., Tzismadia, G., Kab, R., Priboi, C., David, D., Răcătăian, A., Baumgartner, L., Fricker, S., Ruiz, J. F., Armas, E., Brinkhuis, M., & Spruit, M. (2021). A Shared Cyber Threat Intelligence Solution for SMEs. Electronics, 10(23), 2913. https://doi.org/10.3390/electronics10232913

Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K.-R. (Eds.). (2019). Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Lecture Notes in Computer Science, Vol. 11700. Cham: Springer. https://doi.org/10.1007/978-3-030-28954-6

Shneiderman, B. (2022). Human-centered AI. Oxford University Press.

Stoleriu, R., Petre, I., & Pop, F. (2025). Cybersecurity governance in large-scale infrastructures. Romanian Journal of Information Technology and Automatic Control, 35(1), 37–52.

Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. Information Fusion, 76, 89–106. https://doi.org/10.1016/j.inffus.2021.05.009

Weickert, T. D., Joinson, A., & Craggs, B. (2023). Is cybersecurity research missing a trick? Integrating insights from the psychology of habit into research and practice. Computers and Security, 128, Article 103130. https://doi.org/10.1016/j.cose.2023.103130