Páhi Tímea, BSc, MA

The Information Security Challenges of the Digital State

National University of Public Service
Doctoral School of Military Sciences

07. Theory of Defence Informatics and Communication

International Convergence, National Governance, and AI Support in Cyber Exercise Ecosystems



Supervisor: Dr. András Tóth, PhD, Associate Professor

Dokumenteninformation	
Titel	International Convergence, National Governance, and AI Support in Cyber Exercise Ecosystems
Editor	Dr. András Tóth, PhD, Associate Professor
Author	Páhi Tímea, MA, BSc
Description	Doctoral dissertation
Created on	2025/12/05
Last modified on	2026/010/26

Version history			
Version	**Date**	**Change**	**Comments**
0.1	2015/12/09	Modification of table of contents and structure	The right numbering of the chapters, references and table naming will be done at the final steps
0.2	2015/12/28	Add new parts: State-of-the-Art of AI enhanced cyber security and attacker tools (Shannon, PentestGPT, Kali GPT, Penligent etc.)	Creating the scientific methodology for the State-of-the Art analysis and for the Hands-On evaluation (contributions), this part is deleted from the final version
0.3	2026/01/13	Adding description of the national coordination structures of Cyber security in Austria	Creating the chapters about the Austrian CX ecosystem
0.4	2026/01/21	Finalizing Research Questions and Hypothesis, Shifting the scope	Narrowing the scope for raising the quality of the work, deeper technical analysis
0.5	2026/01/25	Adjusting RQs, Hypothesis and Contributions, Re-structuring chapters, Creating cohesion table and presentation	Re-writing, Chapter 1, Deleting parts based on the narrowed scope
0.6	2026/02/09	Re-structuring Chapter 2: From Cyber Exercises through Cyber Ranges to Enhancing National Cyber Resilience	Mering the different chapters with the legal requirements, Summarizing the different types of CXs
0.7	2026/02/10	Research on latest cyber exercises lifecycles, Reading papers after 2024 and AI-enhanced Cyber Exercises	Describing the CX Lifecycle (ENISA), Creating own structure with the AI
0.8	2026/02/12	Creating explaining tables for every chapter, Creating the figures, Re-wrtiting Hypotheses and RQs at least 10times, Setting final point for the structure – no more major changes allowed	Deciding to make a glossary and short explanation brackets for every chapter (like what is the difference between CX and Stress test). Major structural changes could endanger the deadline
0.9	2026/02/14	Research about latest regulation	Design idea – ENISA CX Methodology
0.10	2026/02/16	Re-thinking the contribution after the release of the ENISA CX Methodology	Decision to make: Throwing my categorization and CX lifecycle away and build-up again on ENISA best-practices for better future-references and usage of the dissertation
0.11	2026/02/18	Describing AI-enhanced Cyber Exercise Lifecycle	Finalizing other chapters
0.12	2026/02/19	Describing agentic workflow	Finalizing other chapters, Refining term definition

0.13	2026/02/25-2026/04/22	Creating the figures, Major changes after feedback from supervisor	Creating clear and explanatory figures to the content, add summary text boxes, add personal experience, refining hypotheses
0.14	2026/08/14-17	Finalizing title, figures and references	Implementing the local AI agent
0.15	2026/08/15-19	Proof-reading whole document	Proof reading, adding references in the right style
0.16	2026/08/26-29	Restructuring, new figures	Finalizing clearer structure
0.17	2026/09/28 – 10/19	Finalizing contribution-structure, Re-formulate H1, Describing H3 validation, Explaining use case Austria, Eliminating duplications	Deleting ENISA-Lifecycle as contribution, but as a connecting element, Creating the convergence validation matrix to enhance H1 contribution, Implementing the demonstrator, Executing the validation of the demonstrator and documenting it

Table of Contents

Introduction and research framework ... 12
Research poster and scientific overview ... 12
Relevance and significance of the topic ... 13
Research problem, research gap statement ... 16
Research questions and hypotheses ... 21
Hypothesis H1 ... 22
Hypothesis H2 ... 22
Hypothesis H3 ... 23
Integrated Research Logic of the Dissertation ... 23
Summary of the methodology used ... 26
Structure of the dissertation ... 28
Expected Contributions ... 30
Contribution 1 – European Convergence Framework ... 30
Contribution 2 – National Cyber Exercise Governance Model ... 30
Contribution 3 – Auditable Agentic Workflow ... 30
Cross-Cutting Regulatory and Policy: Regulatory Pathway ... 31
Summary: Research Framework and Three-Layer Structure ... 32
1 Part I – International and European Research Layer ... 33
1.1 Introduction: International Cyber Exercise Landscape, Regulation and Convergence ... 33
1.2 Cyber exercises globally: concepts, typologies, and ecosystem positioning ... 38
Current ecosystem ... 38
Used terminology ... 40
Analytical Dimensions for Comparing Cyber Exercises ... 44
Comparative Mapping of Selected Cyber Exercise Refence Cases ... 46
Summary of International Cyber Exercise Landscape ... 67
1.3 International standards, norms, and best practices for cyber exercises ... 69
Homeland Security Exercise and Evaluation Program (HSEEP) ... 70
ENISA ... 71
BSI Standards ... 74
1.4 EU Strategies and initiatives ... 79
Network and Information Systems Directive (NIS und NIS2) ... 80
EU Cybersecurity Act ... 82
Critical Entities Resilience Directive (CER) ... 82
Digital Operational Resilience Act (DORA) ... 84
Data Protection Law (GDPR) ... 85
EU Cyber Blueprint ... 88
Overview and Summary ... 88
1.5 European Convergence Framework (C1) ... 89
From Fragmentation to Convergence ... 90
Dimensions of the European Convergence Framework ... 91
Convergence in Exercise Purpose: From Training to Resilience Validation ... 93
Convergence in Lifecycle and Artefact Logic ... 93
Regulatory Convergence as a Driver of Exercise Design ... 94
The Emerging Minimum Baseline ... 95
Convergence Does Not Mean Uniformity ... 96
Comparative Evidence Matrix for the European Cyber Exercise Framework (C1) ... 97
Relationship Between the Framework and H1 ... 109
Scientific Contribution and Originality ... 110

From European Convergence to National Governance 113
1.6 Regulatory pathway: convergence trends and mandatory CX 114
1.7 Summarizing H1 International Convergence Hypothesis 115
2 PART II – National research layer 116
2.1 Cyber exercises globally, international convergence and regulatory landscape 116
2.2 Austria as a Reference Case for National Cyber Exercise Governance 120
Austrian Institutional and Strategic Framework 124
National Cybersecurity Coordination Architecture 125
NISG 2026 and the Bundesamt für Cybersicherheit 126
Critical-Entity Resilience: RKEG and RKEV 127
Key Actors in the Austrian Cybersecurity Ecosystem 127
National and Sectoral CSIRT Ecosystem 128
Financial-Sector Resilience and DORA-VG 128
Artificial Intelligence Governance in Austria 128
Austria in the European Cybersecurity Architecture 129
Cyber Exercises in the Austrian Context 129
Analytical Findings from the Austrian Reference Case 129
Austria as the Reference Point for the National Governance Contribution 129
2.3 Cyber Exercise Taxonomy and Minimum Dataset 131
Derivation and Design Logic of the Taxonomy 131
Cyber Exercise Taxonomy 132
Minimum Dataset for National Comparability 134
Integration into National Exercise Governance 137
2.4 National Cyber Exercise Governance Model (C2) and Programme Blueprint 140
National Cyber Strategies – Zero Point 141
Programme objectives and the national capability cycle 144
Governance model and institutional roles 145
Exercise portfolio and progression model 145
Standardized artefacts and a minimum dataset for comparability 146
The national cyber exercise platform as enabling infrastructure 147
Measurement: From exercise completion through KPIs to resilience evidence 148
Institutionalising after-action improvement loops 155
Implementation roadmap: building the programme in phases 156
Scientific Contribution and Added Value of the National CX Governance Model 156
Position of the Programme Concept within the Dissertation Architecture 159
2.5 ENISA-Based AI-Enhanced Cyber Exercise Lifecycle 161
State of the Art of Exercise Lifecycles 162
Selection of ENISA as the Baseline 163
AI Augmentation Principles 164
AI Support Across the Six Lifecycle Phases 165
Human Approval and Governance Gates 171
Evidence and Provenance Flows 174
Methodological Role of the ENISA-Based AI-Enhanced Lifecycle 175
2.6 Regulatory Pathway II – National Governance for Cyber Resilience 176
NIS2 / NISG 2026 177
CER / RKEG and RKEV 177
DORA / DORA-VG 177
GDPR / Austrian Data Protection Framework 177
EU AI Act / Austrian AI Governance 177
Implication for National Cyber Exercise Governance 177
2.7 Assessment of H2 and Part II Summary 177
3 Part III – Organizational and Operational Research Layer 183
3.1 Introduction: From Lifecycle-Level AI Support to Organizational Operationalisation 183

3.2 Auditable Agentic Workflow (C3) ... 183
Concept and Rationale for Agentic Workflow Design ... 184
Design Requirements for the Auditable Agentic Workflow ... 186
Scientific Contribution of Part III ... 203
Research Results and Interpretation ... 203
Validation Framework ... 204
Provenance and Audit Architecture ... 204
Human-in-the-Loop Control Architecture ... 204
Scenario and Inject Engineering ... 205
Supporting Exercise Technologies and Environments ... 208
3.3 Prototype / Demonstrator Architecture ... 212
Empirical Validation of the Agentic AI Cyber Exercise Demonstrator ... 216
Empirical Evaluation and Verification of Hypothesis H3 ... 218
3.4 Regulatory Pathway III – Organizational Compliance Operationalisation ... 219
3.5 Assessment of H3 and Part III Summary ... 221
4 Conclusion and Scientific Contributions ... 221
4.1 Integrated Answer to the Research Problem ... 221
4.2 H1 – International and European Convergence ... 222
4.3 H2 – National Cyber Exercise Governance ... 223
4.4 H3 – Auditable AI Operationalization ... 223
4.5 Validation and Strength of Evidence ... 225
4.6 Limitations ... 225
4.7 Future Research ... 226
4.8 Contributions and Final Conclusion ... 226
5 Declaration of Generative AI Use and Author Responsibility ... 231
6 Abbreviations ... 232
7 Appendix A - Scientific Posters ... 235
8 Appendix B - Cyber Exercises and Formats ... 237
9 References ... 239

List of Figures

FIGURE 1 - SCIENTIFIC POSTER – THE NEXT GENERATION OF CYBER EXERCISES 12
FIGURE 2 - INTERPLAY AMONGST DOMAINS 18
FIGURE 3 - CORE RESEARCH PROBLEM AND 3 RESEARCH GAPS. 19
FIGURE 4 - THREE-LAYER RESEARCH MODEL WITH COMPLIANCE PATHWAYS AND HYPOTHESES 25
FIGURE 5 - METHODOLOGY AND THREE RESEARCH LAYERS 27
FIGURE 6 - STRUCTURE 28
FIGURE 7 - CONTRIBUTION AND REGULATORY PATHWAY SUMMARY 31
FIGURE 8 - INTERNATIONAL / EUROPEAN RESEARCH LAYER OVERVIEW 37
FIGURE 9 - CYBER EXERCISE AND RESILIENCE VALIDATION ECOSYSTEM 41
FIGURE 10 - POSITIONING SELECTED CXS BY SCOPE AND EXERCISE TYPE 47
FIGURE 11 - KYHA NATIONAL CS EXERCISES (SOURCE: JYVSECTEC PRESENTATION) 59
FIGURE 12 - INTERNATIONAL CX LANDSCAPE - SELECTED REFERENCE EXERCISES 68
FIGURE 13 - CONVERGENCE OF EXERCISE METHODOLOGIES 69
FIGURE 14 - HSEEP EXERCISE CYCLE 70
FIGURE 15 - EUROPEAN CONVERGENCE FRAMEWORK FOR CYBER EXERCISES (C1) 93
FIGURE 16 - CONTRIBUTION 1 : CONVERGENCE FRAMEWORK 112
FIGURE 17 - INTERNATIONAL CONVERGENCE TO NATIONAL LAYER 116
FIGURE 18 - OVERVIEW OF NATIONAL RESEARCH LAYER 120
FIGURE 19 - AUSTRIA AS A REFENCE CASE FOR NATIONAL CX GOVERNANCE 123
FIGURE 20 - AUSTRIA CYBERSECURITY GOVERNANCE TRANSITION 126
FIGURE 21 - CX TAXONOMY AND MINIMUM DATASET 130
FIGURE 22 - HUMAN APPROVAL GATES 173
FIGURE 23 - METHODOLOGICAL OVERVIEW: ENISA-BASED AI-ENHANCED LIFECYCLE 176
FIGURE 24 - ENISA-BASED AI-ENHANCED CX LIFECYCLE - CHAPTER OVERVIEW 181
FIGURE 25 - OVERVIEW OF ORGANIZATIONAL RESEARCH LAYER 183
FIGURE 26 - AGENTIC AI WORKFLOW 215
FIGURE 27 - THREE LAYERS AND CONTRIBUTIONS 222
FIGURE 28 – (H1) FROM FRAGMENTED PRACTICE TO A EUROPEAN BASELINE 223
FIGURE 29 - FROM EUROPEAN CONVERGENCE TO NATIONAL CX GOVERNANCE 223
FIGURE 30 - FROM LIFECYCLE-LEVEL AI SUPPORT TO ORGANIZATIONAL IMPLEMENTATION 224
FIGURE 31 - SUMMARY OF CONTRIBUTIONS 228
FIGURE 32 - SCIENTIFIC POSTER – METHODOLOGY 235
FIGURE 33 - SCIENTIFIC POSTER OF FUTURE CX (HU) 236

List of Tables

TABLE 1 - ANALYTICAL DIMENSIONS USED FOR THE COMPARATIVE ASSESSMENT OF CXS 46
TABLE 2 - COMPARATIVE MAPPING OF SELECTED CX REFERENCE CASES 49
TABLE 3 - CYBER EXERCISE TYPES ACCORDING TO BSI 100-4 77
TABLE 4 - EUROPEAN CONVERGENCE FRAMEWORK FOR CXS 92
TABLE 5 - COMPARATIVE EVIDENCE MATRIX: EXERCISE METHODOLOGIES 98
TABLE 6 - COMPARATIVE EVIDENCE MATRIX: EU LEGISLATION 99
TABLE 7 - EVIDENCE CLASSIFICATIONS BY CONVERGENCE DIMENSION 108
TABLE 8 - ASSESSMENT OF REGULATORY CONVERGENCE 108
TABLE 9 - PROPOSED CX TAXONOMY 134
TABLE 10 - MINIMUM CX DATASET 137
TABLE 11 - INTEGRATION OF THE CX TAXONOMY INTO NATIONAL EXERCISE GOVERNANCE 140
TABLE 12 - POSSIBLE KPIS FOR A NATIONAL CYBER EXERCISE PROGRAMME 154
TABLE 13 - AI AUGMENTATION ACROSS THE ENISA LIFECYCLE 171
TABLE 14 - AGENTIC WORKLFOW EXAMPLE ARTEFACT 204
TABLE 15 - IMPROVEMENT COMPARISON 219

Exposé

International Convergence, National Governance, and AI Support in Cyber Exercise Ecosystems

Cyber resilience has become a strategic requirement for modern states because essential public functions, critical infrastructures, defence, finance, healthcare, transport, and communications increasingly depend on interconnected digital systems. Cyber incidents can therefore develop from isolated technical disruptions into complex crises with consequences across organizations, sectors, and national borders. In this environment, **cyber exercises** (later CXs) **provide a controlled way to simulate realistic cyber incidents and to test whether decision-making, communication, escalation, response, recovery, and coordination mechanisms work under pressure**. Their role is consequently expanding from individual training activities toward instruments for validating organizational and national preparedness.
The dissertation addresses the problem that cyber exercises are increasingly expected to be **repeatable, comparable, auditable, and evidence-oriented**, while current practice remains fragmented across countries, sectors, and organizations. Differences can be observed in terminology, exercise types, governance structures, planning methods, documentation, evaluation approaches, and the way lessons learned are translated into measurable improvement. At the same time, European regulations, directives, strategies, and policy initiatives increasingly shape expectations regarding cybersecurity risk management (RM), business continuity management (BCM), critical-entity resilience, incident reporting, third-party dependencies, data protection, and responsible use of artificial intelligence. The research therefore considers compliance and policy not as a separate topic, but as a **top-down pathway from international and European requirements through national governance and legislation to organizational implementation and exercise validation**.

The dissertation is structured around **three interconnected analytical layers**. The **international and European layer** examines standards, methodologies, regulatory drivers, and international cooperation in order to identify fragmentation and emerging convergence in cyber exercise practices. The corresponding **Hypothesis H1** proposes that, although the international and European cyber exercise landscape remains heterogeneous, common standards, European regulation, and policy developments increasingly create a shared baseline for more documented, comparable, auditable, and resilience-oriented exercise practices. The **national layer** then investigates how these external requirements can be translated into a coherent national cyber exercise ecosystem, using Austria as the main reference case. **Hypothesis H2** argues that national cyber exercise governance based on clearly defined institutional roles, common quality criteria, comparable artefacts and datasets, coordinated exercise programmes, and structured improvement loops provides a stronger basis for resilience development than fragmented or ad hoc exercise activity.

Within the national layer, the dissertation develops an **AI-enhanced cyber exercise lifecycle based on the ENISA Cybersecurity Exercise Methodology**. The lifecycle follows the phases of Initiation, Design, Preparation, Execution, Evaluation, and Moving Forward and defines where artificial intelligence can support the exercise process while maintaining human authority, traceability, provenance, bounded AI use, and phase-gated progression. The lifecycle is intended to provide a common methodological bridge between national governance requirements and organizational exercise implementation. It therefore explains **where AI can support the lifecycle and under which governance conditions**, without prescribing a specific technical implementation.

The **organizational and operational layer** subsequently addresses how this AI support can be implemented in practice through an **auditable agentic workflow**. Instead of using AI as an unrestricted general-purpose assistant, the proposed workflow decomposes exercise planning and evaluation into bounded tasks performed by specialised AI-supported components with defined inputs, outputs, responsibilities, and human approval gates. These components can support activities such as requirement analysis, objective formulation, scenario and inject development, evidence structuring, evaluation, reporting, and improvement tracking. **Hypothesis H3** proposes that a controlled and auditable agentic workflow can improve the efficiency, consistency, traceability, and structured

production of cyber exercise artefacts while maintaining human responsibility, legal compliance, security, and accountability.

An important cross-cutting element of the dissertation is the **regulatory and policy pathway connecting the three layers**. European instruments such as the NIS2 Directive, the Critical Entities Resilience Directive, DORA, the GDPR, the EU AI Act, the Cybersecurity Act, and relevant strategic initiatives influence national legislation, governance structures, and competent authorities. These national requirements are subsequently translated into organizational obligations such as cybersecurity risk management, incident reporting, continuity planning, resilience testing, AI governance, human oversight, data protection, supply-chain security, documentation, and improvement tracking. Cyber exercises can then be used to test whether these requirements are not only formally implemented but also function effectively under realistic conditions.

Methodologically, the dissertation applies a **multi-method and constructive research design** combining conceptual and documentary analysis, comparative reference-case analysis, taxonomy and dataset modelling, governance and lifecycle model development, workflow-oriented AI operationalisation, and structured validation. Austria serves as the principal national reference environment for examining how European requirements, national governance structures, competent authorities, sectoral actors, and critical-infrastructure stakeholders interact. The research develops a cyber exercise taxonomy and minimum dataset logic, an implementable national cyber exercise programme blueprint, an ENISA-based AI-enhanced lifecycle, and an auditable agentic workflow. Validation focuses on criteria such as internal consistency, completeness, traceability, governance compatibility, comparability, human control, and the ability to produce structured evidence and support continuous improvement.

The **scientific added value** of the dissertation lies in integrating three areas that are frequently considered separately. First, it links **international and European convergence** in standards, regulation, and exercise methodologies to the development of a common baseline for cyber resilience. Second, it translates this baseline into **national cyber exercise governance**, including institutional coordination, comparable artefacts, lifecycle logic, measurement, and continuous improvement. Third, it operationalises responsible AI support at the **organizational level** through an auditable agentic workflow aligned with the national lifecycle and regulatory requirements. The dissertation therefore proposes a coherent pathway from **international requirements, through national governance, to organizational implementation and evidence generation**, showing how cyber exercises can evolve from isolated training events into scalable, comparable, auditable, and AI-supported instruments for strengthening national cyber resilience.

Exposé

Nemzetközi konvergencia, nemzeti irányítás és mesterséges intelligencia által támogatott működés a kiberbiztonsági gyakorlatok ökoszisztémájában

A kiberreziliencia napjainkra a modern államok egyik alapvető stratégiai követelményévé vált, mivel a közigazgatás, a honvédelem, a pénzügyi rendszer, az egészségügy, a közlekedés és a kritikus infrastruktúrák működése egyre inkább összekapcsolt digitális rendszerekre épül. A kibertámadások következményei ezért túlmutatnak a technikai incidenskezelésen, és szervezeti, ágazati vagy nemzeti szintű válsághelyzeteket idézhetnek elő. A **kiberbiztonsági gyakorlatok** lehetőséget biztosítanak a döntéshozatal, a kommunikáció, az incidenskezelés, a helyreállítás és a szervezetek közötti együttműködés ellenőrzött környezetben történő tesztelésére. Különösen fontos szerepet töltenek be a honvédelmi felkészültség, a katonai–civil koordináció és a kritikus szolgáltatások működőképességének vizsgálatában.

A disszertáció alapvető kutatási problémája, hogy a kiberbiztonsági gyakorlatokkal szemben egyre erősebb követelmény a **megismételhetőség, összehasonlíthatóság, auditálhatóság és a mérhető fejlesztési eredmény**, miközben a terminológia, a gyakorlattípusok, az irányítási modellek és az értékelési eljárások országonként és szervezetenként jelentősen eltérnek. Ezzel párhuzamosan az európai szabályozás – különösen a NIS2, a CER, a DORA, a GDPR és az EU AI Act – egyre konkrétabb elvárásokat támaszt a kockázatkezeléssel, a rezilienciával és a mesterséges intelligencia (MI) felelős alkalmazásával szemben.

A kutatás **három egymásra épülő szintet** vizsgál: a nemzetközi és európai, a nemzeti, valamint a szervezeti és operatív szintet. Ezekhez három hipotézis (H1–H3), három fő kutatási kérdés (MRQ1–MRQ3) és három önálló tudományos hozzájárulás (C1–C3) kapcsolódik.

A **nemzetközi és európai szint** a kiberbiztonsági gyakorlatokra vonatkozó szabványokat, módszertanokat és szabályozási irányokat elemzi. A **H1 hipotézis** szerint a jelenlegi heterogén gyakorlati környezetben a nemzetközi módszertanok és európai követelmények fokozatos konvergenciát eredményeznek a dokumentálhatóbb, összehasonlíthatóbb és auditálhatóbb gyakorlatok irányába. Az első tudományos hozzájárulás, a **C1 – European Convergence Framework**, nyolc elemzési dimenzió mentén rendszerezi a közös követelményeket, és azonosítja a gyakorlatok értékelését támogató európai minimumszint elemeit.

A **nemzeti szint** egy összehangolt kiberbiztonsági gyakorlati ökoszisztéma kialakításának feltételeit vizsgálja, Ausztriát alkalmazva fő referenciaesetként. A **H2 hipotézis** szerint a meghatározott irányítási felelősségekre, közös minőségi kritériumokra, összehasonlítható adatokra és kötelező fejlesztési visszacsatolásra épülő nemzeti rendszer alkalmasabb a rezilienciaorientált képességfejlesztés támogatására, mint az elszigetelt, ad hoc gyakorlatok. A második tudományos hozzájárulás, a **C2 – National Cyber Exercise Governance Model**, egységes gyakorlattipológiát, minimumadatkészletet, teljesítménymutatókat, programirányítási mechanizmusokat és ellenőrzött fejlesztési ciklust foglal magában. A modell különösen a nemzeti koordináció, a katonai–civil együttműködés és a honvédelmi felkészültség értékelése szempontjából hasznosítható.

A nemzeti és szervezeti szint közötti módszertani kapcsolatot az **ENISA módszertanára épülő, MI-támogatott kiberbiztonsági gyakorlat-életciklus** teremti meg. Az Indítás, Tervezés, Előkészítés, Végrehajtás, Értékelés és Továbblépés fázisaiban meghatározza a mesterséges intelligencia lehetséges támogatási pontjait, az emberi jóváhagyási követelményeket és a dokumentálhatóság feltételeit. Az életciklus nem külön tudományos hozzájárulás, hanem a C2 és C3 közötti módszertani összekötő elem.

A **szervezeti és operatív szint** az MI-támogatott gyakorlattervezés megvalósítására és teljesítményének értékelésére összpontosít. A **H3 hipotézis** szerint az ember által felügyelt, RAG-alapú agentic AI munkafolyamat a hagyományos emberi tervezéshez képest javíthatja a gyakorlatok előkészítésének hatékonyságát, minőségét, belső konzisztenciáját és nyomon

követhetőségét. A harmadik tudományos hozzájárulás, a **C3 – Auditable Agentic AI Workflow**, egy ténylegesen megvalósított és empirikusan értékelt háromágenses demonstrátorra épül. Az Objectives & KPIs Agent a mérhető célokat és mutatókat dolgozza ki, a Scenario Architect Agent a forgatókönyveket készíti elő, míg az MSEL / Inject Engineer Agent a strukturált eseménylistát és az injecteket állítja elő. A rendszer szervezeti tudásbázist, visszakereséssel támogatott szöveggenerálást és kötelező emberi jóváhagyási pontokat alkalmaz az ENISA-életciklus első három fázisában.

A H3 empirikus értékelése tényleges kutatási adatokon alapul. A vizsgálat két tanulmányi szemeszterben, négy egyetemi osztályban, összesen 16 hallgatói csoport és nyolc párosított összehasonlítás bevonásával történt. A hagyományos és az MI-támogatott tervezési eljárásokat egységes forgatókönyvi követelmények és értékelési szempontok alapján hasonlították össze. Az MI-támogatott megközelítés **25,3%-kal csökkentette a tervezési ráfordítást**, **7,13 ponttal javította a dokumentumok minőségét** a 100 pontos skálán, **44,6%-kal mérsékelte a belső konzisztenciahibákat**, és **69,6%-ról 88,9%-ra növelte a nyomon követhetőségi lefedettséget**. A kutatást februári, európai uniós szakmai környezethez kapcsolódó nemzetközi résztvevői adatgyűjtés egészítette ki, amely az MI-támogatott gyakorlattervezés szakmai relevanciáját és alkalmazhatóságát vizsgálta. Az eredmények a vizsgált körülmények között empirikusan alátámasztják a H3 hipotézist, általánosíthatóságuk azonban további kutatást igényel.

A disszertáció keresztmetszeti eleme a nemzetközi és európai szabályozási követelmények, a nemzeti irányítás és a szervezeti megvalósítás összekapcsolása. A kiberbiztonsági gyakorlatok így nem csupán képzési események, hanem a szervezeti felkészültség, az együttműködési mechanizmusok és a fejlesztési intézkedések működőképességét vizsgáló eszközök.

A kutatás **hadtudományi hozzáadott értéke** a nemzeti kibervédelmi koordináció, a katonai–civil együttműködés, a kritikus infrastruktúráktól való függőségek és a honvédelmi felkészültség fejlesztésének támogatásában jelenik meg. A C1 összehasonlítható módszertani alapot, a C2 összehangolt gyakorlati irányítási és értékelési keretrendszert, míg a C3 empirikusan vizsgált, emberi felügyelet mellett működő MI-támogatást biztosít.

A disszertáció tudományos hozzáadott értékét a három egymásra épülő eredmény – European Convergence Framework (C1), National Cyber Exercise Governance Model (C2) és Auditable Agentic AI Workflow (C3) – együttesen adja. A kutatás átfogó célja a kiberbiztonsági gyakorlatok megismételhetőségének, összehasonlíthatóságának, auditálhatóságának és mérhető fejlesztési hasznosíthatóságának erősítése, hozzájárulva a nemzeti kiberreziliencia és a honvédelmi felkészültség fejlesztéséhez.

Introduction and research framework

Research poster and scientific overview



FIGURE 1 - SCIENTIFIC POSTER – THE NEXT GENERATION OF CYBER EXERCISES[1]

Relevance and significance of the topic

In an increasingly digitalized world, **cyber resilience has become a core requirement for nation states**. Public administration, defence, emergency response, healthcare, finance, energy, transport, and communications depend on digital infrastructures that are highly interconnected and increasingly exposed to malicious interference.[2] This dependency has two consequences that make the topic of this dissertation particularly relevant. First, cyberattacks have moved from being primarily an organizational IT problem to a systemic risk that can disrupt essential services and critical state functions. Second, the societal and political impact of cyber incidents is amplified by interdependencies: a disruption in one organization can cascade to suppliers, service partners, and downstream operators, and thereby evolve into a broader crisis with national implications. As a result, preparedness is no longer evaluated solely by whether technical controls exist, but by whether organizations and ecosystems can act effectively under pressure and restore essential functions within acceptable timeframes.

In the Hungarian strategic context, Kovács and Krasznay (2024) similarly emphasise that growing dependence on interconnected digital infrastructures creates systemic national vulnerabilities that cannot be addressed solely through technical protection. Their analysis reinforces the need for coordinated preparedness, resilience-oriented capability development, and whole-of-system approaches to cybersecurity.

This development is particularly pronounced in defence-oriented contexts. State and non-state actors increasingly conduct targeted operations against critical infrastructure, communication systems, or military-relevant networks to gain political, economic, or strategic advantage. Responding to such threats requires more than technological upgrades: it requires coordinated strategic and organizational readiness across multiple actors[3][4]. As highlighted in the current dissertation,(military) cyber exercises are central in this context because they enable the simulation of realistic attack scenarios, the testing of defensive measures, and the improvement of coordination between national and international actors[5]. Regular exercising strengthens capabilities, refines processes, and reveals vulnerabilities in an environment where response speed and coordination can determine success or failure. While the defence-oriented framing is essential, the same logic extends to civilian critical services: modern cyber crises often blur traditional boundaries between public and private responsibilities, because essential services are frequently operated by private entities and depend on complex supply chains. Therefore, each nation must maintain a resilient cyberspace, to guarantee robust operation of the nation state[6].

Attacks against critical information infrastructures may also be used as instruments of strategic pressure rather than merely as isolated technical disruptions. Kovács (2023) demonstrates how targeting such infrastructures can generate broader political, economic, and societal effects[7]. This strengthens the case for exercise designs that test not only technical incident response but also

[1] The poster contains AI generated visualizations.
[2] Ristvej, J., Tonhauser, M., Chovanec, D., Kubás, J., Kollár, B., & Zamiar, Z. (2025). Cyber resilience conceptual model for European Union NIS2 standards implementation in Slovakia. Scientific Reports, 15(1), 26902.
[3] Kovács, L. & Krasznay, C. (2024). “Digitális Mohács 3.0.” Hadtudomány, 2024(3), 40–55.
[4] Abazi, B. (2022). Establishing the national cybersecurity (resilience) ecosystem. IFAC-PapersOnLine, 55(39), 42-47.
[5] Páhi, T., & Tóth, A. (2024). The potential of cybersecurity exercises in critical infrastructures. American Journal of Research Education and Development, 2024(3), 13–48.
[6] Ziring, N. (2022, March). National cyber resilience and roles for public and private sector stakeholders. In International Conference on Critical Infrastructure Protection (pp. 3-46). Cham: Springer Nature Switzerland.
[7] Kovács, L. (2023). “Nyomásgyakorlás a kritikus információs infrastruktúrák támadásán keresztül.” In C. Krasznay (Ed.), Taktikák és stratégiák a kiberhadviselésben (pp. 151–168).

continuity, crisis coordination, public communication, and strategic decision-making under conditions of systemic disruption.

The increasing strategic importance of cyberspace in military and national-security operations has also been analysed by Kovács (2023), who positions cyber operations as an integral component of contemporary warfare and capability development[8]. This reinforces the relevance of exercises as mechanisms for preparing organizations and national systems for complex cyber operations that combine technical, operational, and strategic effects.

Cyber exercises are therefore vital because they provide one of the few mechanisms that *can validate preparedness under realistic conditions without waiting for a real incident.* Unlike policy documents, compliance statements, or static risk assessments, exercises *expose how decision-making, communication, escalation, and recovery actually work when time pressure,* uncertainty, and organizational friction are present. Exercises also make socio-technical dynamics visible[9]. Cyber incidents are rarely purely technical; they trigger legal and regulatory obligations, public communication demands, executive trade-offs between service continuity and security controls, and coordination challenges across departments and organizations[10]. This is why exercises have become a cornerstone of preparedness across the full spectrum, from discussion-based tabletop formats *to complex functional and technical simulations*, and why they increasingly serve not only training needs but also assurance needs.

However, the significance of the topic is not limited to the general importance of exercises. It arises from the fact that the ***role of exercises is changing***. Across many contexts, exercises are no longer treated as optional good practice; ***they are becoming expected instruments of resilience governance.*** The consequence is that ***exercises must increasingly meet requirements that go beyond being educational: they must be repeatable, comparable across cycles, and auditable, particularly when they support sectoral resilience targets and national programmes.*** This shift creates a practical and scientific challenge. On the practical side, exercise planners face a growing expectation to produce structured artefacts, collect evidence, and demonstrate improvement over time, often under constraints of limited facilitation capacity and heterogeneous planning maturity. On the scientific side, the field requires approaches that make exercise design and evaluation measurable and reproducible, so that claims about “improved resilience” can be substantiated beyond anecdotal success stories.

The relevance of this dissertation is further reinforced by the growing convergence of regulatory and governance expectations that increasingly shape cyber exercise design, even where exercises are not prescribed explicitly. As resilience frameworks place greater emphasis on continuity, recovery, and dependency management, exercise requirements also evolve. Scenarios can no longer focus exclusively on the internal processes of a single organization; they must incorporate ICT third parties, supply-chain dependencies, and cross-entity coordination mechanisms as integral elements of cyber resilience. In this respect, the dissertation highlights that dependency management, particularly with regard to external ICT providers and interconnected supply chains, is emerging as a cross-cutting requirement across several regulatory drivers. Consequently, *cyber exercises increasingly need to test supplier coordination, escalation pathways, information exchange, and joint recovery processes. This development strengthens the case for national-level*

[8] Kovács, L. (2023). Hadviselés a 21. században: Kiberműveletek. Ludovika Egyetemi Kiadó.

[9] Ristvej, J., Tonhauser, M., Chovanec, D., Kubás, J., Kollár, B., & Zamiar, Z. (2025). Cyber resilience conceptual model for European Union NIS2 standards implementation in Slovakia. Scientific Reports, 15(1), 26902.

[10] Olakojo, K. A., & Virginia, A. (2024). Strengthening cross-border cybersecurity resilience through federated threat intelligence sharing, real-time incident correlation, and coordinated governance mechanisms. Preprint, December. https://doi. org/10.7753/IJCATR1312, 1011, 135.

comparability and coordination, because dependency-related risks and coordination failures cannot be assessed adequately within the organizational boundaries of a single entity. It also underlines the importance of structured learning loops: when similar weaknesses in inter-organizational coordination recur across multiple exercises, resilience governance requires mechanisms to document findings, track remediation, and verify improvement through subsequent re-testing.

At the same time, the dissertation argues that progress in this area remains constrained by a persistent maturity gap in the infrastructures and methodologies required for realistic and scalable cyber exercising. Although cyber ranges have expanded considerably in recent years, there is still no sufficiently standardised methodological basis for their design, configuration, and operation, especially in critical-infrastructure and defence-relevant contexts. This gap is particularly significant in environments characterised by stringent operational requirements, such as high availability, real-time responsiveness, safety constraints, and long system life cycles. The lack of methodological maturity has consequences for both practice and research. From a practical perspective, insufficiently realistic or poorly aligned exercise environments may fail to reflect the operational conditions under which essential systems actually function, thereby limiting the transferability of lessons learned. From a research perspective, the absence of common methodological standards reduces comparability across cases and makes it difficult to assess systematically which exercise designs, technologies, or program structures are most effective in strengthening cyber resilience.

The strategic relevance of the dissertation is further supported by recent expert analyses which argue that artificial intelligence is reshaping the cyber domain not only as a defensive capability, but also as a source of growing attacker advantage. Marsi Tamás notes that cybercriminal actors can integrate AI into offensive operations faster and with fewer regulatory or ethical limitations than defenders, thereby increasing the asymmetry between attack and defence. This development is highly relevant for the present research, because it reinforces the need for preparedness mechanisms capable of testing not only technical responses, but also coordination, information sharing, and collective decision-making under rapidly evolving threat conditions. The same analysis also stresses the growing importance of national and sectoral coordination structures, including CSIRTs, SOCs, ISACs, and AI-supported early-warning ecosystems. From the perspective of this dissertation, these trends confirm that cyber resilience can no longer be treated solely as an organizational issue but must increasingly be approached as a national capability problem requiring structured and repeatable exercise-based validation.[11]

A further reason why the topic is relevant these days is the rapid diffusion of artificial intelligence into cyber security practice, including the planning and execution of exercises. AI is frequently presented as a solution to capacity problems: it can draft artefacts faster, help generate scenario variants, assist in monitoring inject delivery, and support evaluation through faster evidence processing. Yet unstructured use of AI creates new risks, particularly in regulated and defence-oriented settings where accountability and traceability are not negotiable. The dissertation therefore treats AI not as a replacement for human judgement, but as a capability that must be integrated through controlled processes. The dissertation captures this principle explicitly through the “AI recommends, humans decide” logic and by stressing the need for full decision logging and version control across the ENISA-based exercise lifecycle.

The relevance of artificial intelligence to defence and national security is also reflected in recent Hungarian military-scientific research. The edited volume A mesterséges intelligencia hatásainak

[11] Marsi, T. (2025, December 17). 2026 a kiberreziliencia éve. ICT Global Magazin. https://ictglobal.hu/print/kiberbiztonsag-mi-2026-marsi-tamas/

specifikus vizsgálat by Kenedli and Magyar[12] addresses the broader implications of AI across technological, organisational, legal, and national-security domains. This multidisciplinary perspective reinforces the importance of examining AI not merely as a technical capability, but as a technology whose adoption must be considered within established defence responsibilities, institutional structures, and security requirements. The present dissertation extends this research context by investigating the controlled application of agentic AI to cyber exercise preparation and evidence-oriented resilience development.

Recent Hungarian defence and national-security research **has emphasised the strategic importance of integrating artificial intelligence into military and security-related organisational processes,** while maintaining human responsibility, ethical safeguards, and practical applicability[13]. The research discussions reported by Magyar and Péter also highlight the importance of pilot implementations, operational evaluation, and AI-supported organisational learning. These considerations provide a relevant military-scientific context for the present dissertation, which develops and empirically evaluates an auditable, human-supervised agentic AI workflow for cyber exercise preparation.

This leads to the dissertation's core significance: ***it addresses not only whether cyber exercises are important, but how they can be scaled and governed as a national resilience instrument without compromising accountability***. The work is positioned at the intersection of three needs that are simultaneously increasing: (i) the strategic need for nation-state cyber resilience and defence-oriented preparedness; (ii) the governance need for repeatable, comparable, and auditable exercises that feed improvement loops; and (iii) the operational need to reduce the labour intensity and variability of exercise planning and evaluation. Within this space, the dissertation *introduces the concept of auditable, phase-gated AI support, operationalized through an agentic workflow, to strengthen artefact consistency and traceability across the exercise lifecycle*. The dissertation motivates this by noting that agentic workflows fit the inherently modular and sequential nature of exercise planning and allow controlled go/no-go checkpoints that make outputs reviewable and defensible.

In summary, the topic of this dissertation is relevant because cyber resilience is now a state-level imperative, and cyber exercises are one of the most effective mechanisms to operationalise preparedness across technical, organizational, and inter-organizational dimensions. It is significant because the expectations placed on exercises are evolving toward national-scale comparability, auditability, and demonstrable improvement, requirements that current practice often struggles to fulfil consistently. ***By addressing governance, lifecycle design, and controlled AI augmentation within a coherent framework, the dissertation targets a problem that is both practically urgent and scientifically underexplored, with direct implications for how nation states can build and evidence cyber resilience over time***.

Research problem, research gap statement

The rise of the digital state has made cyber resilience a strategic requirement for national governance, since essential public functions now rely on interconnected digital infrastructures, data flows, and service ecosystems. As Tóth András's concept of the digital state implies, the secure

[12] Kenedli, T., & Magyar, S. (Eds.). (2025). A mesterséges intelligencia hatásainak specifikus vizsgálata. Katonai Nemzetbiztonsági Szolgálat. https://knbsz.gov.hu/kiadvanyok

[13] Magyar, S., & Péter, K. F. (2025). A mesterséges intelligencia hatásainak specifikus vizsgálata című tudományos–szakmai konferencia és könyvbemutató. Szakmai Szemle, 23(3–4). https://www.knbsz.gov.hu/ac57d5f9-b367-4a2a-a604-5e7309961a6d

and reliable operation of this ecosystem is no longer a purely technical issue, but a prerequisite for state continuity, public trust, and effective governance.

According to Tóth András, the digital state is not limited to electronic public administration but represents a broader state-level digital ecosystem that includes internal governmental IT, citizen- and business-facing e-government services, and other public digital services such as health, education, libraries, cultural heritage, and public data-sharing functions. Its effective operation requires not only the provision of services, but also the security background necessary to ensure their reliable and protected functioning. Tóth identifies the main building blocks of the digital state as robust digital infrastructure, adequate digital competences across society and public administration, and a strong digital economy that supports both service provision and innovation. In functional terms, the digital state depends on broadband availability, continuous service accessibility, the uptake and use of digital technologies, interoperable public services, and the capacity to create public value in a data-driven manner while maintaining trust and privacy protection. This interpretation is directly relevant to the present dissertation, because it shows that state resilience in the digital era depends not only on technological modernisation, but also on the secure, coordinated, and trustworthy operation of a complex socio-technical public ecosystem.[14]

This dependency has created a structural asymmetry: while digital transformation increases efficiency and connectivity, it also expands attack surfaces and produces tightly coupled interdependencies that can amplify local technical failures into national-scale disruption. In this environment, cyberattacks are not only a matter of information security; they can undermine state capacity to act, degrade essential services, and erode trust in institutions. For defence-oriented contexts in particular, cyber operations increasingly form part of broader strategic competition, with state and non-state actors targeting critical infrastructure, communication systems, and military networks to gain political, economic, or strategic advantage.

Cyber exercises are one of the few instruments that allow states and critical service operators to confront these realities in a controlled setting. Exercises make it possible to rehearse decision-making under uncertainty, validate coordination routines across organizational boundaries, and test whether detection, response, and recovery capabilities function under stress. They also provide a mechanism to identify systemic weaknesses before they become national incidents. In other words, cyber exercises are vital because they offer an operational bridge between abstract risk management and proven readiness. This importance is consistent with NATO's understanding of cyberspace as a domain of operations and of cyber defence as part of the Alliance's core task of deterrence and defence. ***NATO explicitly treats exercises as essential tools for testing and validating concepts, procedures, systems, tactics, interoperability, and crisis coordination,*** while cyber-specific exercises such as Cyber Coalition or Locked Shields exercises are used to strengthen the ability of NATO, Allies, and partners to protect networks, operate in cyberspace, exchange information in real time, and improve collective cyber resilience.[15] From this perspective, cyber exercises are not merely training events, but strategic mechanisms for readiness validation, operational learning, and collective defence in an increasingly contested digital environment. Yet despite the growing number of exercise activities and supporting environments, including cyber ranges, the national-level preparedness effect is still difficult to demonstrate in a consistent and defensible way[16].

[14] András, T. (2022). A digitális állam információbiztonsági kihívásai. Budapest: Ludovika.
[15] North Atlantic Treaty Organization. (2024, July 30). Cyber defence. https://www.nato.int/en/what-we-do/deterrence-and-defence/cyber-defence
[16] Tabansky, L., & Lichterman, E. (2025). PROGRESS: the sectoral approach to cyber resilience. International Journal of Information Security, 24(1), 18.

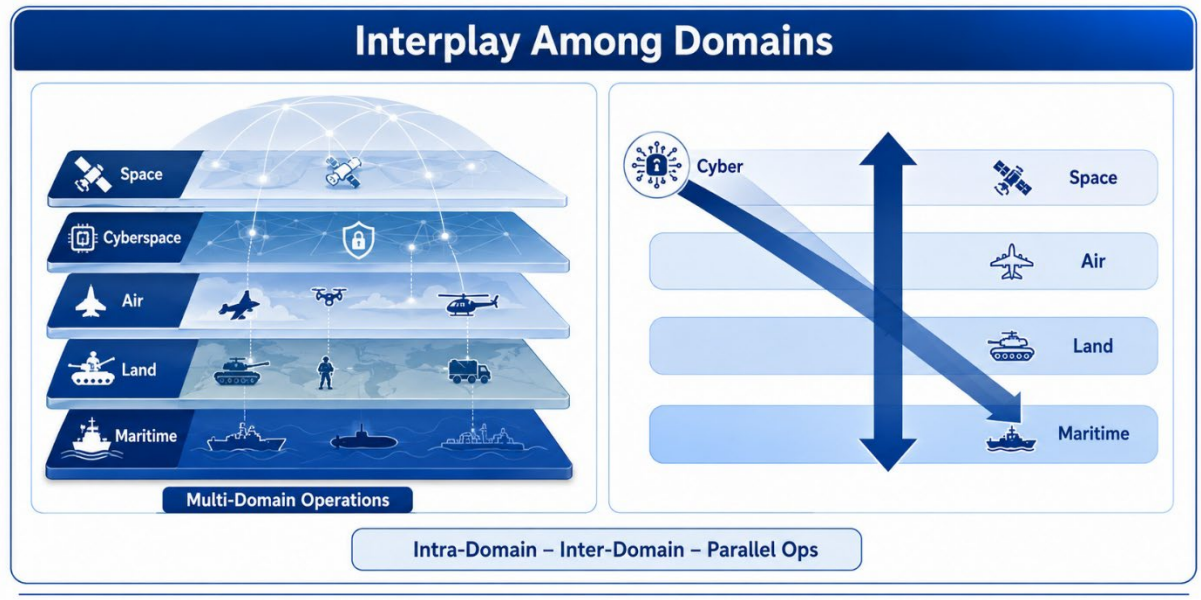


FIGURE 2 - INTERPLAY AMONGST DOMAINS[17]

Figure 2 illustrates the framework of Multi-Domain Operations across space, cyberspace, air, land, and maritime environments. It maps out their interactions through intra-domain, inter-domain, and parallel operations while showing how cyberspace connects every physical layer.

The core research problem addressed in this dissertation is therefore formulated as follows:

Cyber exercises are increasingly expected to strengthen and provide evidence of national cyber resilience, but current practice lacks the integrated governance, comparability, improvement mechanisms, and controlled AI support needed to make cyber exercising repeatable, auditable, and scalable.

The research problem is therefore not the absence of cyber exercises. Cyber exercises already exist at organizational, sectoral, national, European, and international levels. Rather, the problem lies in the absence of a sufficiently integrated system through which heterogeneous exercise activities can be compared, their findings translated into verified improvement, and their planning and evaluation responsibly scaled through artificial intelligence.

Three interrelated research gaps arise from this central problem: **comparability, improvement, and responsible scaling**.

[17] Created based on Laudun, J., Kroh, T., Sidikki, M., Arp, R., & Lowther, A. (2021, 21. Oktober). The Department of Defense's multidomain operations challenge. Global Security Review. https://globalsecurityreview.com/defense-department-multidomain-operations-challenge/

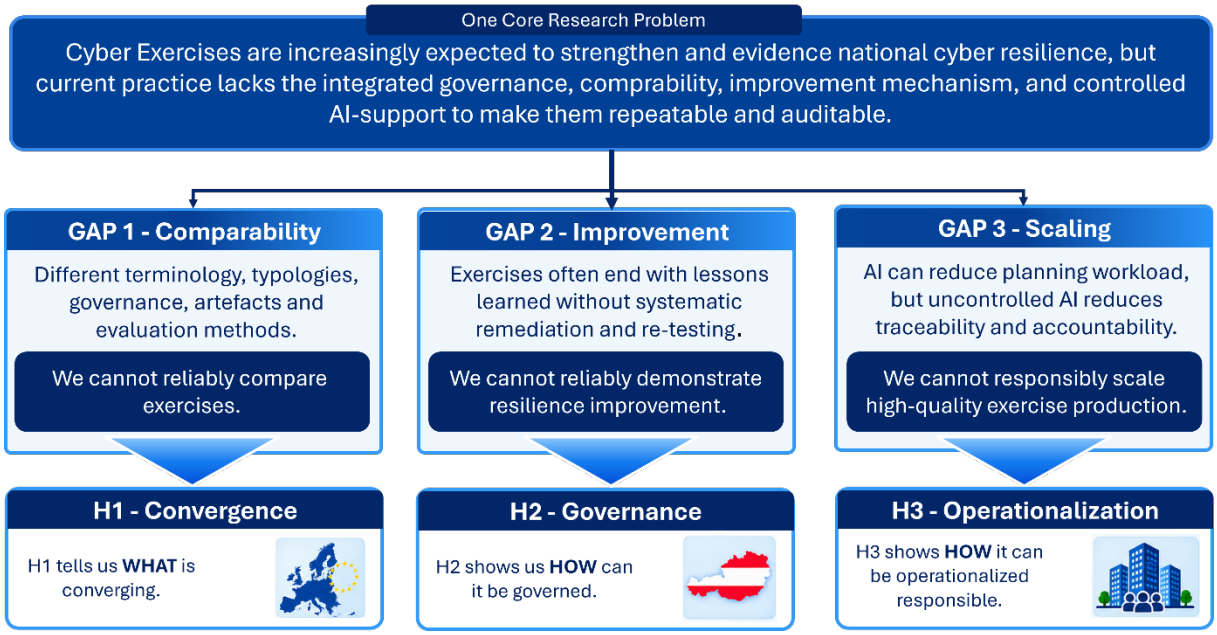


FIGURE 3 - CORE RESEARCH PROBLEM AND 3 RESEARCH GAPS.

Figure 3 summarises the central research problem and shows how the three identified gaps provide the conceptual basis for the three research layers, hypotheses, and scientific contributions of the dissertation.The figure illustrates how the central research problem is decomposed into three interrelated gaps - comparability, improvement, and responsible scaling - which subsequently provide the conceptual basis for H1, H2, and H3 and for the three scientific contributions of the dissertation.

GAP 1 - Comparability: The first dimension of this research problem concerns comparability. Cyber exercises are conducted using different terminology, typologies, methodologies, governance arrangements, artefact structures, and evaluation approaches. Differences also exist in organizational maturity, planning resources, facilitation capability, documentation quality, and exercise objectives. Consequently, the fact that two organizations or countries both conduct a “cyber exercise” does not necessarily mean that they test comparable capabilities or produce comparable evidence. This fragmentation becomes particularly problematic when exercise results are considered beyond the level of a single organization. National and European stakeholders may need to understand what capabilities were tested, under which assumptions, through which methods, and on the basis of what evidence. If exercise programmes use incompatible structures and reporting practices, meaningful comparison remains difficult. Without common conceptual principles, structured artefacts, minimum datasets, and shared evaluation logic, national stakeholders cannot reliably aggregate or compare results across organizations, sectors, or exercise cycles. In practical terms, this means that cyber exercises cannot yet be reliably compared. This comparability gap provides the basis for the international and European research layer and for H1, which examines whether international methodologies, standards, and European regulatory developments are creating convergence towards a more comparable, auditable, and resilience-oriented cyber exercise baseline.

GAP 2 - Improvement: The second dimension concerns improvement. Cyber exercises frequently conclude with an after-action review, lessons identified, recommendations, or an After-Action Report. Yet the identification of a weakness does not itself demonstrate that resilience has improved. Improvement becomes observable only when findings are assigned to responsible actors, translated into corrective actions, implemented, monitored, and subsequently re-tested. Without this closed improvement loop, organizations may repeatedly identify the same weaknesses across successive exercises without demonstrating meaningful capability development. This problem is particularly relevant in national and defence-oriented contexts because many critical

weaknesses occur at the interfaces between organizations rather than within individual technical systems. Examples include escalation thresholds, coordination between authorities and operators, crisis communication, third-party dependencies, continuity prioritisation, and cross-sector response mechanisms. The resilience value of an exercise therefore lies not merely in detecting weaknesses but in connecting exercise evidence to a continuous sequence of findings, corrective actions, re-testing, and verified improvement. In conceptual terms, this logic can be expressed as **Exercise → Evidence → Finding → Corrective Action → Re-test → Verified Improvement**. The absence of such institutionalised loops means that resilience improvement cannot yet be demonstrated reliably. This gap provides the basis for the national research layer and H2, which investigates whether a structured national cyber exercise governance model can provide a stronger foundation for comparability, accountability, and resilience-oriented improvement than isolated or ad hoc exercise activities.

GAP 3 - Scaling: The third dimension concerns responsible scaling. High-quality cyber exercise planning is resource-intensive and requires the analysis of requirements, definition of objectives, development of realistic scenarios, preparation of injects, coordination of stakeholders, identification of evidence sources, evaluation planning, reporting, and follow-up activities. Artificial intelligence offers considerable potential to support several of these tasks. It may assist in synthesising information, generating scenario alternatives, drafting exercise artefacts, checking consistency across outputs, supporting evaluation, and accelerating reporting. At the same time, uncontrolled use of generative AI introduces new risks. AI-generated artefacts may contain undocumented assumptions, inconsistent information, unsupported claims, hallucinations, or unclear source provenance. Such weaknesses are particularly problematic in defence-oriented and regulated environments, where exercise artefacts and planning decisions must remain reviewable, attributable, explainable, and auditable. The relevant research question is therefore not simply whether AI can make exercise production faster, but whether AI can be embedded into the exercise process in a manner that preserves human authority, governance control, traceability, provenance, and quality assurance. Without such safeguards, high-quality cyber exercise production cannot be responsibly scaled through AI. This challenge forms the basis of the organizational and operational research layer and H3, which investigates how AI support can be operationalized through a bounded, auditable, and human-controlled Agentic Workflow.

The three gaps are closely interconnected and together define the central research challenge of the dissertation. The comparability gap limits the establishment of a consistent evidence base across exercises, organizations, sectors, and countries. The improvement gap limits the transformation of that evidence into demonstrable capability development. The responsible scaling gap limits the ability to expand and accelerate high-quality exercise production without introducing new risks to accountability, traceability, and trust. The research gap addressed by this dissertation is therefore not the lack of cyber exercises as such, but the absence of an **integrated, evidence-driven cyber exercise system** capable of linking common principles, governance, exercise execution, evidence, corrective action, re-testing, and responsible AI support.

This integrated problem directly structures the three research layers and the three principal scientific contributions of the dissertation. The international and European layer addresses the comparability problem by examining fragmentation and emerging convergence and by developing the **European Convergence Framework for Cyber Exercises**. This first contribution identifies common principles and emerging minimum expectations that can provide a more comparable and auditable baseline for cyber exercising. The national layer addresses the governance and improvement problem by developing the **National Cyber Exercise Governance Model**, including a programme blueprint, common taxonomy, minimum dataset, KPI and improvement framework, and an ENISA-based AI-enhanced cyber exercise lifecycle. The organizational and operational layer addresses the responsible scaling problem by developing the **Auditable Agentic Workflow**

for Cyber Exercises, through which AI support is operationalized through structured inputs and outputs, defined responsibilities, provenance, human approval gates, and controlled execution.

The resulting research logic is cumulative rather than parallel. H1 examines **what is converging** at international and European level. H2 examines **how this emerging baseline can be translated into structured national governance and continuous capability development**. H3 examines **how AI support can be operationalized responsibly within this governed exercise system**. A cross-cutting regulatory and policy pathway connects the three levels by analysing how international and European expectations influence national governance structures and are subsequently translated into organizational responsibilities, exercise objectives, evidence requirements, and improvement mechanisms. The overall objective is therefore to transform cyber exercises from isolated training activities into **repeatable, comparable, auditable, evidence-oriented, continuously improving, and responsibly AI-supported instruments of national cyber resilience**.

Research questions and hypotheses

This dissertation addresses a central problem in defence-oriented and critical-infrastructure cyber preparedness: although cyber exercises, cyber ranges, and resilience-testing initiatives have become increasingly common, they do not yet consistently function as a comparable, auditable, and evidence-oriented instrument of national cyber resilience. Current practice remains fragmented across countries, sectors, and organizations. Exercise terminology, typologies, governance structures, planning methods, artefacts, evaluation approaches, and improvement mechanisms differ substantially, making it difficult to compare exercises, aggregate findings, and demonstrate resilience improvement over time.

At the same time, the environment in which cyber exercises are conducted is becoming increasingly shaped by European regulation, national cybersecurity governance, resilience requirements, and emerging rules for the responsible use of artificial intelligence. Cyber exercises therefore need to be understood not only as training events, but also as instruments through which organizations and national authorities can test whether cybersecurity, continuity, crisis-management, coordination, and regulatory requirements function effectively under realistic conditions.

To reflect this complexity, the dissertation is structured around **three connected analytical layers: international and European, national, and organizational-operational**.
A cross-cutting **regulatory and policy pathway** connects these layers by examining how international and European requirements influence national legislation and governance and how these requirements are subsequently translated into organizational responsibilities, controls, and exercise objectives (see the Figure 4).

The overall research objective is to develop a scientifically grounded and practically applicable framework that connects international convergence, national cyber exercise governance, and auditable AI-supported organizational implementation.

More specifically, the dissertation seeks to:
(1) establish a structured international and European baseline of cyber exercise standards, methodologies, regulatory drivers, and emerging convergence trends;

(2) develop an implementable national cyber exercise governance model, including a common programme logic, taxonomy, comparable artefacts and datasets, improvement mechanisms, and an ENISA-based AI-enhanced cyber exercise lifecycle; and

(3) operationalise responsible AI support at organizational level through an auditable, phase-gated agentic workflow that preserves human control, traceability, and accountability.

These objectives are addressed through three hypotheses and three corresponding main research questions.

Hypothesis H1

The first hypothesis addresses the **International fragmentation and emerging European convergence.**

H1: The international and European cyber exercise landscape remains fragmented in terminology, exercise typologies, governance structures, lifecycle models, and evaluation practices. At the same time, international best practices, European regulatory requirements, and policy developments create convergence towards more documented, comparable, auditable, and resilience-oriented cyber exercise practices.
H1 is primarily an analytical research proposition. It examines whether common principles and emerging minimum expectations can be identified despite the continued diversity of national and sectoral approaches.

This leads to the first main research question:

> **MRQ1:** Which common methodological principles and regulatory requirements can be identified across international and European cyber exercise approaches, and how can these be consolidated into a common framework for comparability and auditability?

The expected contribution associated with H1 is a consolidated analytical baseline of international methodologies, standards, regulatory drivers, and emerging convergence trends. This baseline also forms the starting point of the dissertation's regulatory and policy pathway by identifying which requirements subsequently influence national governance and organizational implementation.

Hypothesis H2

The second hypothesis addresses the **National cyber exercise governance as an enabling framework for resilience-oriented improvement**.

H2: A national defence-oriented cyber exercise ecosystem based on defined governance responsibilities, common quality criteria, comparable artefacts and datasets, coordinated exercise programmes, a structured lifecycle, and mandatory improvement loops provides a more suitable framework for comparability, auditability, and resilience-oriented improvement than fragmented or ad hoc exercise activity. The hypothesis reflects the central governance argument of the dissertation: conducting more exercises does not automatically improve national cyber resilience. Their value depends on whether individual exercises are embedded within a common capability system that allows findings, evidence, and improvement measures to be compared and followed across organizations and exercise cycles.
Austria is used as the principal reference case for examining how European requirements, national cybersecurity governance, responsible authorities, sectoral structures, and exercise activities interact in practice.

This leads to the second main research question:

> **MRQ2:** How can a national defence-oriented cyber exercise ecosystem be designed and governed so that it translates international and European requirements into comparable, auditable, and continuously improving national cyber exercise practice?
> The expected contributions associated with H2 include an implementable national cyber exercise programme blueprint, a governance and coordination model, a cyber exercise taxonomy, a minimum dataset, a KPI-oriented measurement framework, and structured after-action improvement mechanisms.

In addition, the national layer incorporates an **AI-enhanced cyber exercise lifecycle based on the ENISA Cybersecurity Exercise Methodology**. The lifecycle provides a common methodological structure across the phases of Initiation, Design, Preparation, Execution, Evaluation, and Moving Forward. The dissertation extends this structure by defining where AI can support individual lifecycle activities and under which governance conditions such support should operate. Core principles include human authority, traceability and provenance, bounded AI use, and phase-gated progression. The lifecycle therefore serves as the methodological bridge between national governance and organizational implementation.

Hypothesis H3

The third hypothesis addresses the organizational and **Auditable AI operationalisation at organizational level**.

H3: A human-supervised, RAG-grounded agentic AI workflow enables more efficient, internally consistent, traceable, and higher-quality cyber exercise preparation than conventional human-only planning, when evaluated under comparable scenario requirements and standardised assessment criteria.

The dissertation therefore distinguishes between the **AI-enhanced lifecycle** and the **Agentic Workflow**. The lifecycle defines *where* AI support can be introduced within the exercise process and under which governance principles. The Agentic Workflow defines *how* this support can be implemented operationally. Instead of treating AI as a general-purpose assistant, the proposed workflow decomposes exercise planning and evaluation into bounded tasks carried out by specialised AI-supported components. These functions support activities such as requirement analysis, scope definition, objective and KPI development, scenario construction, MSEL and inject engineering, evidence organisation, evaluation, and reporting. Each component receives defined inputs, produces structured outputs, and operates under explicit human approval.

This leads to the third main research question:

> **MRQ3:** To what extent does a human-supervised agentic AI workflow improve the efficiency, artefact quality, internal consistency, and traceability of defence-oriented cyber exercise preparation compared with conventional human-only planning?

The expected contribution associated with H3 is an auditable agentic workflow with specialised responsibilities, structured artefacts, provenance-aware outputs, explicit human approval gates, and controlled progression between workflow stages.

Integrated Research Logic of the Dissertation

A parallel **regulatory and policy pathway** connects all three layers: Three-Layers and the regulatory and policy pathway. European regulations, directives, methodologies, and policy initiatives create requirements and expectations at international and European level. These are translated through national legislation, strategies, governance structures, competent authorities, and sectoral arrangements. At organizational level, they become concrete requirements concerning areas such as cybersecurity risk management, incident handling, business continuity, critical-entity resilience, third-party dependencies, data governance, human oversight, AI governance, documentation, testing, and improvement.

Cyber exercises provide an important mechanism for testing whether these requirements function effectively in practice rather than existing only as formal policies or compliance statements.

The central scientific proposition of the dissertation is therefore that national cyber resilience becomes more governable and improvable when **international requirements, national governance, organizational implementation, and exercise-based evidence are connected within one coherent system**.

The originality of the dissertation lies in integrating three areas that are often examined separately: **international and European regulatory convergence, national cyber exercise governance, and auditable AI-supported organizational operationalisation**.

Across all three layers, a regulatory and policy pathway connects European and international requirements with Austrian national governance and, finally, with concrete organizational responsibilities and exercise objectives.

The Figure 4 below presents the overall research architecture of the dissertation across three interconnected analytical layers: the international/European, national, and organizational/operational levels. At each layer, a specific research problem is linked to a corresponding main research question, hypothesis, and scientific contribution. H1 examines international fragmentation and emerging European convergence and leads to the European Convergence Framework; H2 translates this baseline into a structured national cyber exercise governance model; and H3 operationalizes responsible AI support through an auditable, human-controlled Agentic Workflow. The cross-cutting regulatory and policy pathway connects all three layers by showing how EU regulation, international standards, legal and ethical requirements, auditability, accountability, interoperability, and data governance influence the research architecture throughout. Overall, the figure illustrates the cumulative logic of the dissertation: from identifying what is converging, through determining how it can be governed nationally, to showing how it can be operationalized responsibly at organizational level.

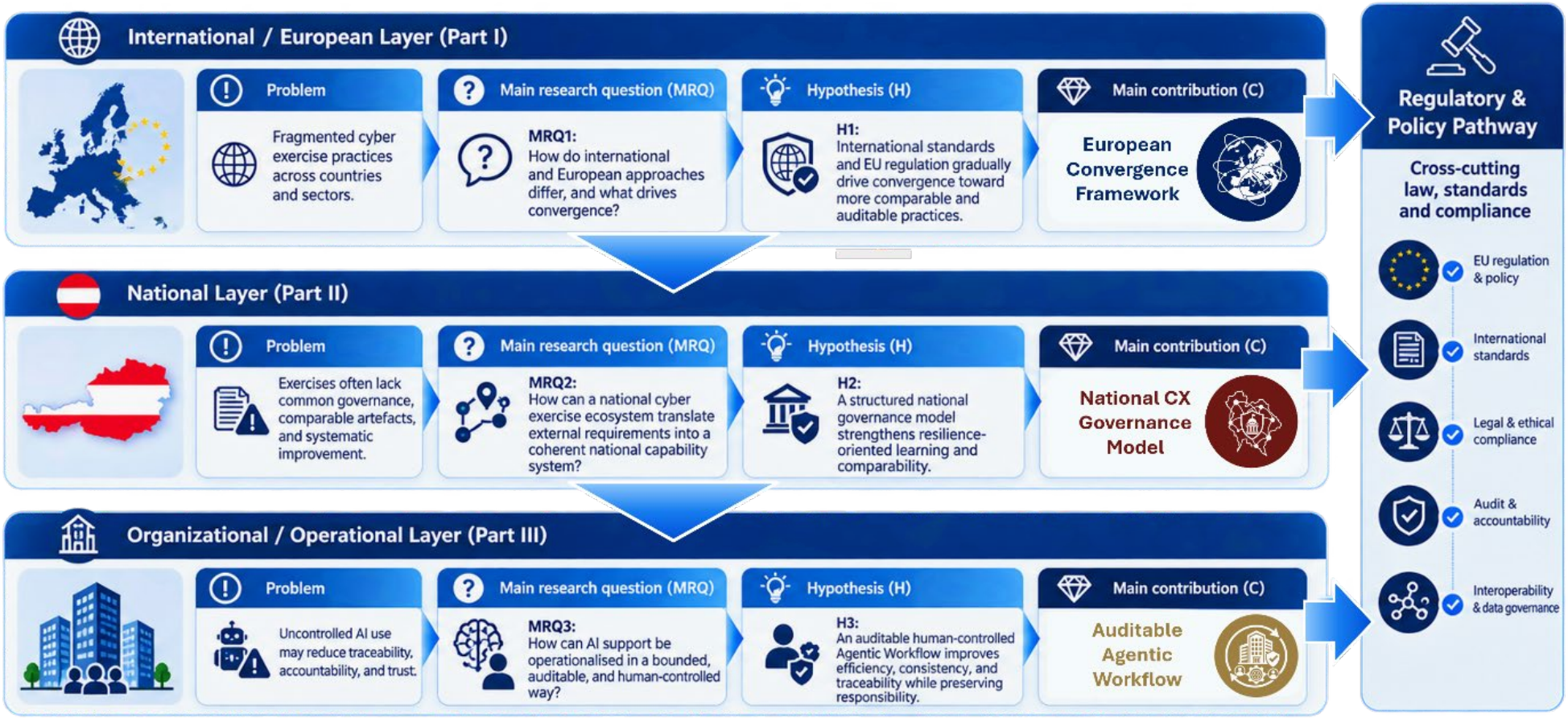


FIGURE 4 - THREE-LAYER RESEARCH MODEL WITH COMPLIANCE PATHWAYS AND HYPOTHESES

Summary of the methodology used

This dissertation applies a multi-method research design combining analytical and constructive research. This methodological approach follows directly from the three-layer structure of the research problem. The dissertation does not investigate a single isolated phenomenon; instead, it examines how international and European developments influence national cyber exercise governance and how these governance requirements can subsequently be translated into organizational practice and AI-supported exercise processes (see in Figure 5).

The methodological design therefore follows the same **cumulative logic as the research framework**: international and European analysis → national model development → organizational operationalisation and validation.

International and European Layer

The first methodological component is **qualitative conceptual and documentary analysis**. This component primarily supports H1 and MRQ1.
International standards, cyber exercise methodologies, regulatory instruments, policy documents, and relevant academic literature are analysed to identify differences and common principles in terminology, lifecycle structures, exercise typologies, evaluation approaches, governance expectations, and resilience requirements.
Particular attention is given to the relationship between cyber exercise methodology and European regulatory or policy developments. The analysis therefore not only asks how exercises are currently conducted, but also which external requirements increasingly influence their purpose, documentation, evaluation, and expected outcomes.
This component produces the international and European baseline on which the later national model is constructed.

National Layer

The second methodological component is **comparative and reference-case-based analysis**, with Austria serving as the principal national reference environment.
The purpose is not to develop an exclusively Austrian model, but to use Austria as a realistic governance setting in which the translation from European requirements to national structures can be examined. This includes institutional roles, national legislation and strategies, coordination mechanisms, sectoral actors, information sensitivity, regulatory responsibilities, and existing exercise activities.
The third component is **taxonomy and dataset modelling**. The dissertation develops a cyber exercise taxonomy and minimum dataset that allow heterogeneous exercises to be described through common attributes. This provides a methodological basis for comparability across exercise types, organizations, sectors, and cycles.
The fourth component is **constructive governance and lifecycle model development**. At national level, the dissertation develops an implementable cyber exercise programme blueprint containing governance structures, programme objectives, exercise portfolio logic, common artefacts, comparable datasets, KPI-oriented measurement mechanisms, and structured improvement loops.
Within the same national layer, the research develops an **AI-enhanced cyber exercise lifecycle based on the ENISA methodology**. Rather than creating an independent lifecycle, the dissertation uses ENISA as the normative European baseline and extends it with clearer AI augmentation points and governance conditions.

The lifecycle defines four cross-cutting principles for AI integration:

- human authority;
- traceability and provenance;
- bounded AI use; and
- phase-gated progression.

For each ENISA lifecycle phase, the dissertation identifies the purpose of the phase, potential AI enhancement, required human control, and expected outputs.

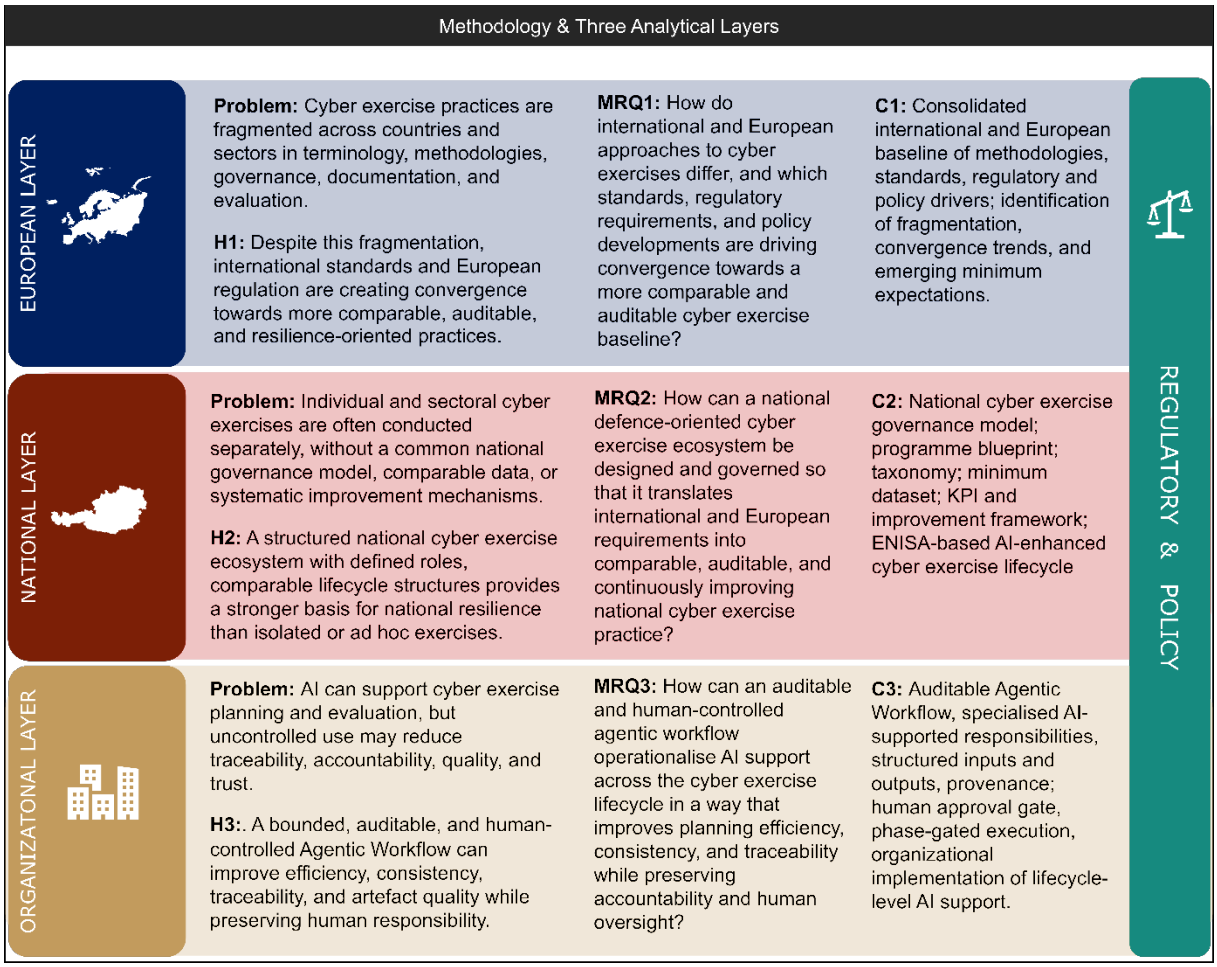


FIGURE 5 - METHODOLOGY AND THREE RESEARCH LAYERS

Organizational and Operational Layer

The fifth methodological component is **workflow-oriented AI operationalisation** and primarily addresses H3 and MRQ3.
The Agentic Workflow is treated both as a practical implementation model and as a research instrument. It translates the lifecycle-level AI augmentation points into specialised and bounded AI-supported tasks.
Each workflow component receives defined inputs and produces structured outputs corresponding to recognised exercise artefacts. The workflow operates sequentially, and phase transitions are controlled through explicit go/no-go criteria and human approval.
This structure enables the research to examine AI support at the level of observable work products rather than relying on general claims about AI usefulness.

Structured Validation

The sixth methodological component is **structured validation**.
Validation focuses on criteria directly related to the research problem, including:

- completeness;
- internal consistency;
- traceability;
- provenance;
- governance compatibility;
- comparability;
- human control;

- regulatory and policy alignment;
- structured evidence generation; and
- support for improvement tracking.

For the Agentic Workflow, validation can additionally examine whether relationships between artefacts remain consistent - for example, whether objectives are linked to scenarios and injects, whether injects generate observable evidence, and whether findings can be traced back to defined exercise objectives.
The dissertation does not claim universal empirical proof that the proposed models increase national cyber resilience in every context. Instead, it evaluates whether the models are coherent, applicable, reproducible, auditable, and compatible with the governance and accountability requirements identified through the earlier analytical stages.

Methodologically, the dissertation can therefore be understood as a progression from **analysis to model construction and from model construction to controlled operationalisation**.

H1 is addressed primarily through conceptual, documentary, regulatory, and comparative analysis.
H2 is addressed through reference-case analysis, taxonomy development, governance modelling, lifecycle development, and comparability mechanisms.
H3 is addressed through workflow design, operationalisation, and artefact-based validation.

Together, these methods create a coherent research design connecting **regulation, governance, exercise methodology, AI operationalisation, and evidence generation**.

Structure of the dissertation

The dissertation follows the same three-layer structure as the research model and hypotheses and a cross-cutting regulatory pathway (see Figure 6 below). It moves from the international and European environment, through national governance and lifecycle design, to organizational and operational implementation.

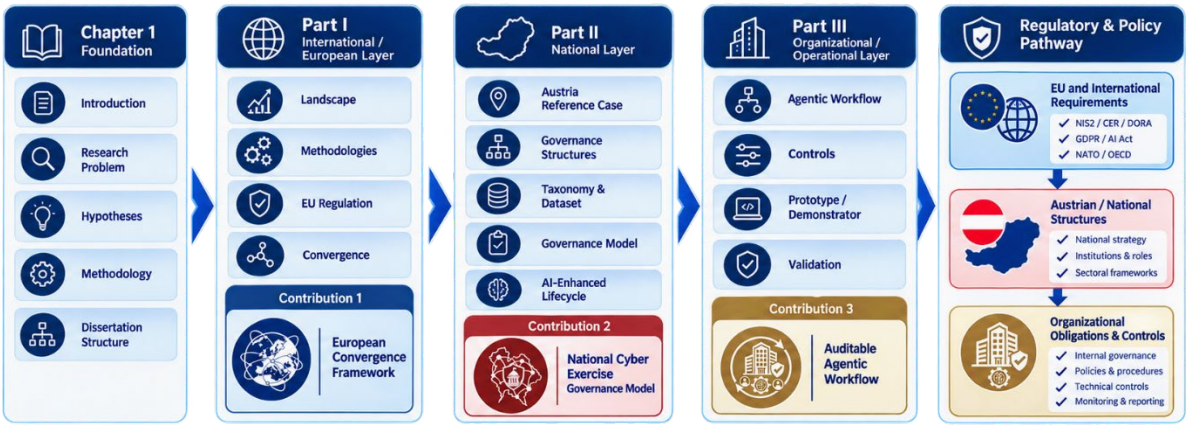


FIGURE 6 - STRUCTURE

Introduction and Research Framework (Chapter 1)

The introductory chapter establishes the relevance of cyber resilience and explains the role of cyber exercises as controlled mechanisms for testing preparedness, decision-making, response, recovery, and coordination.
It defines research problems, research gaps, hypotheses, research questions, methodology, dissertation structure, and expected scientific contributions. It also introduces the three analytical

levels - international, national, and organizational - and the cross-cutting regulatory and policy pathway connecting them.

Part I – International and European Landscape, Regulation, and Convergence

Part I corresponds primarily to **H1 and MRQ1**.
It establishes the international and European baseline for the dissertation. The chapter defines cyber exercises and related terminology, reviews major international methodologies and best practices, and analyses differences in exercise structures, maturity, and evaluation approaches.
A central component of this part is the analysis of European regulatory and policy drivers. Rather than treating compliance requirements as an isolated legal topic, the dissertation examines how they influence expectations concerning preparedness, continuity, cybersecurity risk management, critical-entity resilience, incident response, documentation, third-party dependencies, evidence generation, and responsible AI.

Part I concludes by identifying emerging convergence trends and deriving an international and European baseline for the subsequent national model.

Part II – National Cyber Exercise Governance and AI-Enhanced Lifecycle

Part II corresponds primarily to **H2 and MRQ2**.
It translates the international and European baseline into a national cyber exercise governance model, with Austria serving as the principal reference case. The chapter analyses the national institutional framework, key actors, competent authorities, relevant national legal and policy structures, and the way international and European requirements are translated into national governance.
It then examines existing national and international exercise formats and develops a cyber exercise taxonomy and minimum dataset to improve comparability.
Based on these findings, the dissertation proposes an implementable national cyber exercise programme blueprint covering governance responsibilities, programme objectives, exercise portfolio and progression, standardised artefacts, minimum datasets, measurement, reporting, and continuous improvement.

The national layer also incorporates the **ENISA-based AI-enhanced Cyber Exercise Lifecycle**.
The lifecycle provides a common methodological structure for Initiation, Design, Preparation, Execution, Evaluation, and Moving Forward and identifies where AI can support these phases while maintaining human authority, traceability, bounded use, and controlled progression.
This lifecycle acts as the methodological bridge between national governance and organizational implementation.

Part III – Organizational and Operational AI Implementation

Part III corresponds primarily to **H3 and MRQ3**.
This part operationalises the AI-enhanced lifecycle through an **auditable Agentic Workflow**.
The workflow translates lifecycle-level AI augmentation into specialised and bounded AI-supported responsibilities. It defines the required inputs, expected outputs, dependencies, human approval points, provenance requirements, and go/no-go mechanisms between workflow stages.
The organizational layer demonstrates how requirements originating at European and national level can become concrete controls and exercise artefacts within an organization. Supporting sections address scenario and inject engineering, evaluation artefacts, relevant exercise technologies, evidence collection, and implementation considerations.
The central purpose of this part is therefore to demonstrate how responsible AI support can move from a conceptual lifecycle principle to a reproducible and auditable organizational process.

Conclusion and Contributions
The final chapter consolidates the findings across all three analytical levels and explicitly evaluates H1, H2, and H3.

It explains how international convergence creates the baseline for national governance, how national governance structures enable comparable and improvement-oriented cyber exercising, and how organizational AI-supported workflows can operationalise these requirements.

The chapter also distinguishes between established principles adopted from literature and regulation, original contributions developed in the dissertation, validated properties of the proposed models, limitations, and future research opportunities.

Expected Contributions

The dissertation contributes to the field through three mutually dependent scientific contributions aligned with the three research layers and hypotheses. The originality lies in connecting international and European convergence, national cyber exercise governance, and organizational AI-supported implementation within one evidence-oriented resilience model.

Contribution 1 – European Convergence Framework

The first contribution is a **European Convergence Framework** for cyber exercises. It consolidates international methodologies, standards, recurring exercise practices, European regulatory requirements, and policy developments into a common analytical structure. The framework identifies persistent fragmentation while deriving eight convergence dimensions: terminology and conceptual clarity; governance and accountability; lifecycle orientation; structured artefacts and documentation; evaluation and measurement; auditability and evidence; dependency and ecosystem management; and continuous improvement.
The contribution does not claim completed European harmonization. Instead, it defines an emerging minimum baseline that allows heterogeneous exercises to become more comparable and auditable without requiring identical scenarios, technologies, or exercise formats. This provides the principal analytical response to H1 and MRQ1.

Contribution 2 – National Cyber Exercise Governance Model

The second contribution is an implementable **National Cyber Exercise Governance Model**. Using Austria as the principal reference case, it translates European and international expectations into a national programme architecture based on defined governance responsibilities, an exercise portfolio, common artefacts, a cyber exercise taxonomy, a minimum dataset, KPI-oriented measurement, structured reporting, corrective-action ownership, and re-testing.

Contribution 3 – Auditable Agentic Workflow

The third scientific contribution (C3) is the development and implementation of an **Auditable Agentic AI Workflow** for cyber exercise preparation. Its practical operationalisation is demonstrated through a functioning three-agent system comprising the Objectives & KPIs Agent, Scenario Architect Agent, and MSEL / Inject Engineer Agent, supported by retrieval-augmented generation, structured exercise artefacts, and mandatory human approval gates. The demonstrator establishes the technical feasibility of implementing bounded AI-supported exercise preparation within the Initiation, Design, and Preparation phases of the ENISA methodology, while providing the foundation for comparative quantitative evaluation against conventional human-only planning.
The novelty lies not in using generative AI to draft exercise content, but in transforming AI assistance into a reproducible, reviewable, and governance-compatible process suitable for

regulated and defence-oriented environments. This provides the operational response to H3 and MRQ3.

The ENISA-based AI-Enhanced Cyber Exercise Lifecycle serves as a methodological bridge between the National Cyber Exercise Governance Model (C2) and the Auditable Agentic AI Workflow (C3). It is not classified as a separate scientific contribution.

Cross-Cutting Regulatory and Policy: Regulatory Pathway

A regulatory and policy pathway connects all three contributions. International and European requirements create convergence pressure; national legislation, strategies, authorities, and sectoral arrangements translate those requirements into governance; organizations then operationalize them through controls, responsibilities, exercise objectives, evidence, and improvement actions. Cyber exercises are therefore treated as practical validation mechanisms through which formal requirements can be tested under realistic conditions.
The integrated scientific proposition is: international and European convergence establishes the external baseline; national governance transforms that baseline into a comparable capability system; and organizational AI-supported workflows operationalise the system while preserving accountability. Exercise evidence and improvement loops then support repeatable and auditable national cyber resilience development.

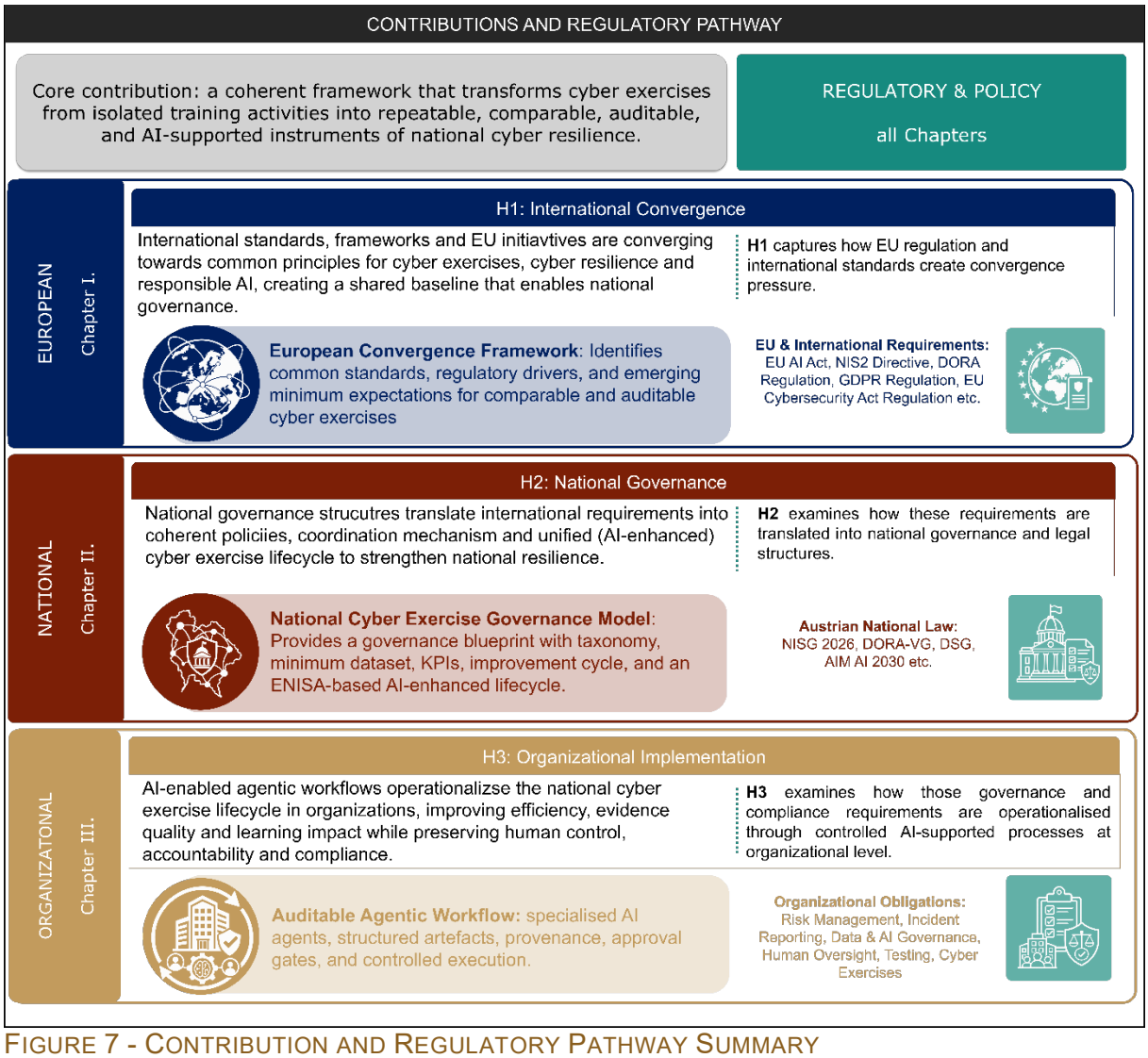


FIGURE 7 - CONTRIBUTION AND REGULATORY PATHWAY SUMMARY

Summary: Research Framework and Three-Layer Structure

This dissertation examines how cyber exercises can strengthen national cyber resilience by testing how organizations detect, manage, communicate, and recover from serious cyber incidents. The research starts from the observation that cyber exercises are becoming increasingly important, while their terminology, design, governance, documentation, and evaluation still differ significantly between countries and organizations. To address this fragmentation, the dissertation is structured into three connected layers: ***international and European, national, and organizational/operational.***

The **international and European layer** analyses standards, methodologies, regulatory requirements, and cooperation mechanisms that influence how cyber exercises are designed and assessed. **Hypothesis H1** argues that, despite continuing fragmentation, international standards and European regulation are gradually creating convergence towards more comparable, auditable, and resilience-oriented exercise practices.

The **national layer** examines how these external requirements can be translated into a coherent national cyber exercise ecosystem, using Austria as the main reference case. It considers governance structures, competent authorities, sectoral actors, critical infrastructures, common quality criteria, comparable artefacts, and continuous improvement mechanisms. **Hypothesis H2** proposes that a structured national cyber exercise system provides a stronger basis for resilience-oriented improvement than isolated or ad hoc exercise activities. At this level, the dissertation also introduces an **AI-enhanced cyber exercise lifecycle based on the ENISA methodology**, covering Initiation, Design, Preparation, Execution, Evaluation, and Moving Forward. The lifecycle defines where AI can support exercise activities while maintaining human authority, traceability, provenance, bounded use, and explicit approval between phases.

The **organizational and operational layer** focuses on the practical implementation of this AI support through an auditable **Agentic Workflow**. Specialised AI components support tasks such as requirement analysis, objective definition, scenario and inject development, evidence structuring, evaluation, reporting, and improvement tracking, while final decision-making remains with human experts. **Hypothesis H3** argues that such a controlled and auditable AI-supported workflow can improve the efficiency, consistency, traceability, and quality of cyber exercise planning without weakening accountability or legal and security requirements.

A cross-cutting **regulatory and policy pathway** connects all three layers by showing how international and European requirements are translated into national governance and subsequently into concrete organizational obligations and exercise objectives.
The cross-cutting regulatory pathway is embedded across H1–H3:

- **H1** captures how EU regulation and international standards create convergence pressure.
- **H2** examines how these requirements are translated into national governance and legal structures.
- **H3** examines how those governance and compliance requirements are operationalized through controlled AI-supported processes at organizational level.

The scientific added value of the dissertation lies in integrating **international convergence, national cyber exercise governance, and auditable AI-supported operationalization** into one coherent framework. In this way, cyber exercises are positioned not only as training events, but as repeatable, comparable, evidence-oriented instruments for strengthening national cyber resilience.

1 Part I – International and European Research Layer

1.1 Introduction: International Cyber Exercise Landscape, Regulation and Convergence

Cyber exercises have become an increasingly important instrument for strengthening preparedness, validating response capabilities, and supporting cyber resilience across organizational, sectoral, national, and international levels. Their role has developed considerably beyond traditional training. In contemporary cyber-resilience environments, exercises are increasingly used to test whether organizations and public authorities are able to detect, assess, communicate, coordinate, respond to, and recover from complex cyber incidents under realistic conditions. This development is especially relevant where cyber incidents may affect essential services, critical infrastructures, defence capabilities, public administration, or interconnected supply chains and therefore have consequences extending beyond the boundaries of a single organization.

This shift reflects the broader research problem introduced in Chapter 0. Cyber exercises are increasingly expected to contribute to the strengthening and evidencing of national cyber resilience, yet current exercise practice remains heterogeneous in terminology, methodology, governance, artefacts, evaluation, and improvement mechanisms. As a consequence, the existence of numerous exercises does not in itself create a comparable or auditable evidence base for resilience. The first analytical task of the dissertation is therefore to determine whether, despite this fragmentation, a sufficiently coherent international and European baseline is emerging from established exercise methodologies, standards, regulatory requirements, and policy developments.

Part I addresses this task through the **international and European research layer** and provides the analytical foundation for Hypothesis H1. H1 proposes that the international and European cyber exercise landscape remains fragmented, while international best practices, European regulation, and policy developments increasingly create convergence towards more documented, comparable, auditable, and resilience-oriented exercise practices. The objective of Part I is consequently not to demonstrate that European cyber exercise practice has already become harmonized. Instead, the analysis seeks to identify where diversity persists, which principles are recurring across different approaches, and whether these recurring elements can be synthesized into an emerging minimum baseline for cyber exercising.

A central conceptual distinction is important in this context. Convergence does not imply uniformity. Cyber exercises differ legitimately because they serve different purposes, sectors, threat environments, organizational maturity levels, and governance structures. A national crisis-management tabletop exercise cannot and should not be identical to a live-fire technical defence exercise, a threat-led penetration test, or an international military cyber exercise. The relevant question is therefore not whether all exercises should use identical scenarios, technologies, or execution formats. Rather, the question is whether common principles can be identified concerning how exercises are governed, planned, documented, evaluated, and connected to improvement. This distinction forms the analytical basis of the **European Convergence Framework for Cyber Exercises** developed later in this part.

The starting point for this analysis is the international cyber exercise ecosystem. Cyber exercises exist within a broader environment of crisis-management exercises, continuity tests, resilience assessments, technical testing, cyber ranges, red teaming, adversary emulation, and threat-led penetration testing. Although these activities overlap in certain respects, they validate different capabilities and therefore cannot be treated as interchangeable. Part I first establishes a clear

conceptual and terminological foundation for distinguishing these forms of resilience validation. This is particularly important for comparability, because meaningful analysis depends on identifying what an exercise is intended to test, at which level it operates, and what evidence its design can reasonably generate.

Chapter **1.2, Cyber Exercises Globally: Concepts, Typologies, and Ecosystem Positioning**, therefore develops the conceptual basis for the subsequent analysis. It positions cyber exercises within the broader resilience-validation ecosystem and distinguishes between discussion-based, functional, distributed, technical, live-fire, and adversarial testing formats. It also clarifies the role of enabling environments such as cyber ranges, simulations, emulations, and digital twins. The chapter further examines selected recurring international and national exercise programmes in order to demonstrate the diversity of contemporary cyber exercising. Examples include European exercises such as Cyber Europe and Blue OLEx, NATO and NATO CCDCOE exercises such as Cyber Coalition, Locked Shields, and Crossed Swords, national programmes such as Germany's LÜKEX and Finland's KYHA exercises, as well as additional national and sectoral reference cases. The purpose of this analysis is not to rank these exercises, but to understand which dimensions are necessary for describing them consistently.

This analysis also supports the development of a common categorisation logic. Because the term *cyber exercise* is applied to highly different activities, comparison requires a structured representation of characteristics such as exercise purpose, organiser, attribution level, sector, exercise type, assessed capability level, scenario, duration, participant scale, and recurrence. Part I therefore establishes the empirical and conceptual basis for the cyber exercise taxonomy and minimum dataset developed more constructively in Part II. In this sense, the international landscape analysis serves two purposes: it demonstrates fragmentation and simultaneously identifies the attributes through which otherwise heterogeneous exercises can become analytically comparable.

Chapter **1.3, International Standards, Norms, and Best Practices for Cyber Exercises**, examines whether established methodologies already contain common principles capable of reducing this fragmentation. The analysis focuses particularly on recognised approaches such as the U.S. Homeland Security Exercise and Evaluation Program (HSEEP), ENISA guidance and methodologies, and the German BSI standards and exercise concepts. Although these frameworks differ in institutional origin, scope, terminology, and intended users, they demonstrate considerable overlap in their underlying exercise logic. Recurring elements include objective-driven planning, clearly assigned roles, structured exercise lifecycles, documented scenarios and injects, evaluation planning, evidence collection, after-action reporting, corrective actions, and continuous improvement.

This methodological comparison is important for H1 because it allows the dissertation to distinguish between visible diversity in exercise formats and deeper convergence in quality principles. Different frameworks may organise their lifecycle phases differently or use different labels for similar artefacts, yet they increasingly rely on comparable assumptions regarding preparedness and evidence. Exercises are expected to have explicit objectives, a defined scope, structured preparation, controlled execution, planned evaluation, and a mechanism for converting findings into improvement. These principles provide one component of the emerging convergence baseline developed later in Part I.

A particularly important contemporary development is ENISA's 2026 *Cybersecurity Exercise Methodology*. Its end-to-end lifecycle of **Initiation, Design, Preparation, Execution, Evaluation, and Moving Forward** provides a current European methodological reference for planning, conducting, evaluating, and following up cyber exercises. Part I analyses this methodology as part of the international and European convergence process, while Part II later adopts it as the recognised baseline for the dissertation's AI-enhanced cyber exercise lifecycle. This separation is

methodologically important: Part I analyses ENISA as evidence of methodological convergence, whereas Part II uses the methodology constructively within the proposed national governance model.

Methodological convergence alone, however, does not fully explain the changing role of cyber exercises. The second major driver is the evolution of the European regulatory and policy environment. Chapter **1.4, EU Strategies and Initiatives: European Fragmentation, Maturity Assessment, and Regulatory Drivers**, therefore analyses the main European instruments that increasingly influence expectations regarding preparedness and resilience. These include the NIS2 Directive, the EU Cybersecurity Act, the Critical Entities Resilience Directive, the Digital Operational Resilience Act, the GDPR, the EU Cyber Blueprint, and other relevant initiatives. These instruments do not all regulate cyber exercises in the same manner. Some establish explicit testing requirements for specific sectors, while others indirectly shape exercise requirements through obligations concerning cybersecurity risk management, incident preparedness, continuity, recovery, governance accountability, third-party dependencies, data protection, or crisis coordination.

The dissertation therefore does not argue that all European legislation mandates cyber exercises. Such a statement would be legally and analytically inaccurate. Rather, the research examines how these instruments collectively produce **convergence pressure** towards more structured resilience validation. Where organizations are required to demonstrate that risk-management arrangements, continuity processes, escalation procedures, reporting mechanisms, or third-party dependencies function effectively, exercises become one practical method for generating evidence about operational preparedness. In specific areas, such as DORA and threat-led penetration testing, testing requirements are considerably more explicit. In other areas, including NIS2 or critical-entity resilience, exercises serve as a suitable means of validating capabilities that are required by the broader governance framework even when a specific exercise format is not prescribed.

This regulatory analysis is also linked to the cross-cutting **regulatory and policy pathway** of the dissertation. At the international and European layer, regulation, standards, and policy create external requirements and expectations. At the national layer, these are translated through legislation, competent authorities, strategies, institutional responsibilities, and national governance structures. At organizational level, they become operational responsibilities, controls, evidence requirements, testing activities, and improvement measures. Part I therefore identifies the origin of this pathway and establishes the regulatory and policy expectations that Part II subsequently examines through the Austrian reference case and the National Cyber Exercise Governance Model.

The analysis of international methodologies and European regulatory drivers converges in Chapter **1.5, European Convergence Framework for Cyber Exercises**, which constitutes the principal scientific contribution of Part I. The framework synthesises the recurring principles identified throughout the international landscape, methodological comparison, and regulatory analysis into eight convergence dimensions: **terminology and conceptual clarity; governance and accountability; lifecycle orientation; structured artefacts and documentation; evaluation and measurement; auditability and evidence; dependency and ecosystem management; and continuous improvement**.

These dimensions represent the dissertation's interpretation of an emerging European minimum baseline. They do not prescribe one universal exercise methodology or one mandatory exercise type. Rather, they establish common characteristics through which heterogeneous exercises can become more comparable and auditable while remaining context-sensitive. A tabletop exercise and a technical live-fire exercise may therefore remain fundamentally different in execution while still being assessed through common questions: Were objectives clearly defined? Were governance

and responsibilities explicit? Were artefacts structured? Was evidence collected? Were findings evaluated systematically? Were corrective actions assigned? Was improvement subsequently verified?

Chapter **1.6, Regulatory Pathway: Convergence Trends and Implications for Mandatory Cyber Exercise Baselines**, further examines how these convergence dimensions are reinforced by the European regulatory environment. Particular attention is given to the distinction between regulation-driven resilience expectations and formally mandatory exercise requirements. The chapter considers whether developments in European cyber-resilience governance indicate a gradual movement from voluntary exercising towards more formalised and evidence-oriented testing expectations. The analysis remains deliberately cautious: convergence towards common preparedness and evidence requirements does not necessarily imply the future introduction of one uniform mandatory European cyber exercise regime. Instead, the emerging trend is towards minimum expectations concerning governance, testing, documentation, evidence, accountability, and improvement that may be realised through different exercise and resilience-testing mechanisms depending on the sector and context.

Part I concludes in Chapter **1.7** by assessing **H1 – the International Convergence Hypothesis**. The hypothesis is evaluated against the evidence collected from the international exercise landscape, comparative methodology analysis, and European regulatory and policy developments. The expected finding is not full harmonisation but evidence of **fragmentation combined with emerging convergence**. Persistent differences remain in terminology, exercise portfolios, institutional structures, sectoral maturity, and implementation practices. At the same time, common principles increasingly appear across methodologies and regulatory expectations, supporting the proposition that a more comparable, auditable, evidence-oriented, and resilience-driven baseline is emerging.

The scientific function of Part I is therefore to establish **what is converging**. Its contribution is analytical: it transforms a heterogeneous landscape of exercise methodologies, regulatory instruments, and practices into a structured European convergence baseline. This baseline subsequently becomes the external reference for Part II, where the research question changes from identifying common expectations to determining **how these expectations can be translated into a coherent national cyber exercise governance system**.

The transition from Part I to Part II is consequently central to the cumulative logic of the dissertation. The European Convergence Framework does not itself create national cyber resilience. It identifies the common principles and minimum expectations that national governance must translate into institutional roles, programme structures, common artefacts, comparable datasets, lifecycle requirements, measurement mechanisms, and continuous improvement. Part II therefore takes the analytical result of H1 as its starting point and investigates its implementation through the national research layer, using Austria as the principal reference case.

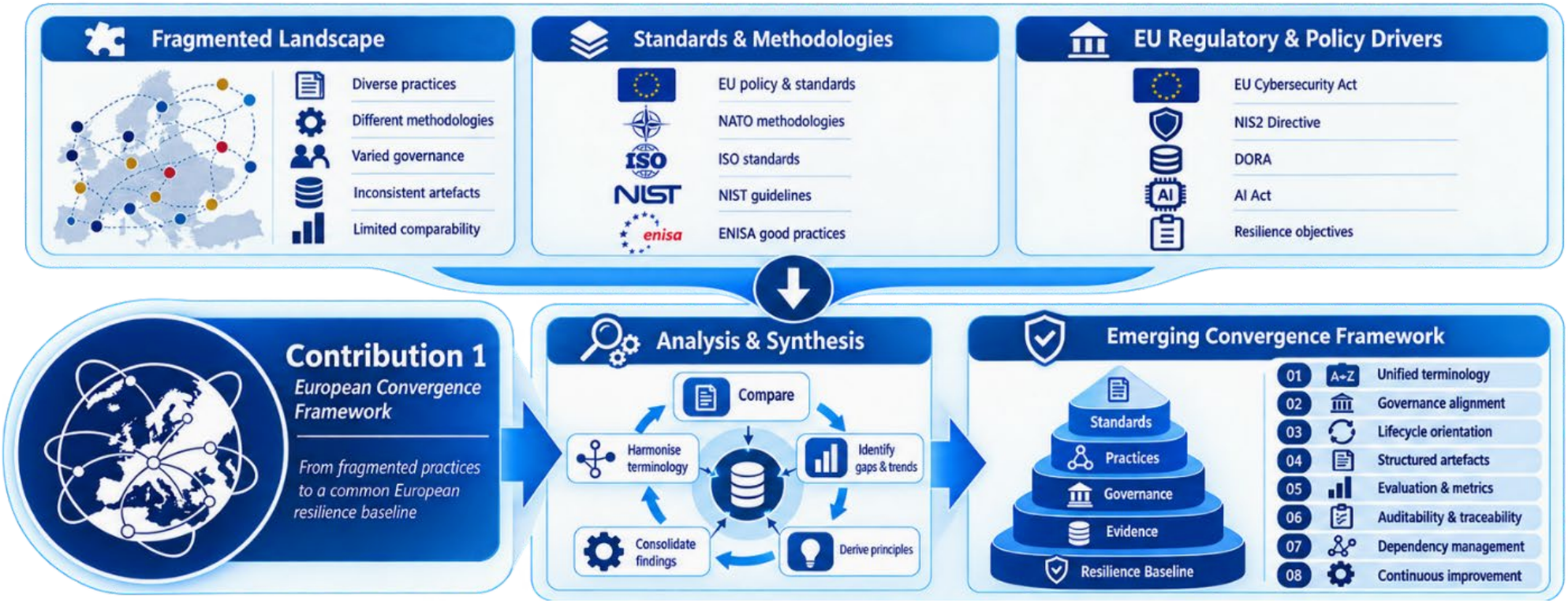


FIGURE 8 - INTERNATIONAL / EUROPEAN RESEARCH LAYER OVERVIEW

Figure 8 presents the analytical logic of Part I. It shows how the investigation progresses from the heterogeneous international cyber exercise landscape through established methodologies and European regulatory and policy drivers towards analysis and synthesis in the European Convergence Framework. The framework establishes the external baseline for the National Cyber Exercise Governance Model developed in Part II.

1.2 Cyber exercises globally: concepts, typologies, and ecosystem positioning

Cyber exercises cannot be understood adequately as isolated training events. They are embedded in a broader ecosystem of institutions, infrastructures, methods, and governance arrangements that jointly shape what can be exercised, what can be observed, and what can be improved. This ecosystem includes public authorities, defence institutions, regulators, operators of essential services, cyber range providers, academia, private security companies, and sector-specific coordination bodies. It also includes the artefacts and rules through which exercises are structured: scenarios, injects, evaluation plans, communication protocols, reporting formats, and improvement tracking mechanisms. The practical importance of this ecosystem is that cyber resilience is rarely produced by one actor in isolation. It emerges from coordinated action across technical, organizational, and inter-organizational layers.

From a scientific perspective, ecosystem positioning matters because cyber exercises differ fundamentally in purpose and design. Some are discussion-based and primarily validate leadership decision-making, communication, and escalation logic. Others are functional exercises focused on process execution, coordination, and role clarity under time pressure. Still others are technical or live-fire simulations that test detection, containment, and recovery in cyber range environments. These differences are not secondary details; they determine what kinds of conclusions can be drawn from the exercise and whether results can be compared across contexts. A national executive tabletop, a sectoral crisis-management staff exercise, and a live adversary-emulation exercise may all be “cyber exercises,” yet they validate different capability layers and require different evaluation approaches.

Cyber exercises cannot be understood adequately as isolated training events. They are embedded in a broader ecosystem of institutions, infrastructures, methods, and governance arrangements that jointly shape what can be exercised, what can be observed, and what can be improved. This ecosystem includes public authorities, defence institutions, regulators, operators of essential services, cyber range providers, academia, private security companies, and sector-specific coordination bodies. It also includes the artefacts and rules through which exercises are structured: scenarios, injects, evaluation plans, communication protocols, reporting formats, and improvement tracking mechanisms. The practical importance of this ecosystem is that cyber resilience is rarely produced by one actor in isolation. It emerges from coordinated action across technical, organizational, and inter-organizational layers.

From a scientific perspective, ecosystem positioning matters because cyber exercises differ fundamentally in purpose and design. Some are discussion-based and primarily validate leadership decision-making, communication, and escalation logic. Others are functional exercises focused on process execution, coordination, and role clarity under time pressure. Still others are technical or live-fire simulations that test detection, containment, and recovery in cyber range environments. These differences are not secondary details; they determine what kinds of conclusions can be drawn from the exercise and whether results can be compared across contexts. A national executive tabletop, a sectoral crisis-management staff exercise, and a live adversary-emulation exercise may all be “cyber exercises,” yet they validate different capability layers and require different evaluation approaches.

Current ecosystem

Cyber exercises have become a central measure for strengthening emergency and crisis management preparedness in organizations. This development reflects a broader shift: cyber

incidents are no longer treated as purely technical disruptions, but as complex socio-technical crises that can affect business continuity, regulatory compliance, public trust, and ,in the case of critical infrastructure ,the delivery of essential services. Consequently, there is now a wide variety of cyber exercises conducted at national, European, and international levels, complemented by private and sector-driven initiatives. Across these formats, the shared purpose is to simulate difficult, complex, and realistic situations ,such as cyberattacks, security incidents, extortion attempts, or compound disruptions ,in order to build readiness for the real case and improve the efficiency and reliability of organizational response.

A core value of cyber exercises lies in their practical impact on preparedness and performance. By rehearsing escalation paths, coordination routines, and operational procedures in a controlled setting, organizations can improve their preparation and response time and thereby shorten and reduce the consequences of cyber incidents . Importantly, exercises can be designed and executed at different scales, making them suitable for diverse organizational contexts and objectives ,from small team-level drills to large, multi-organization simulations that involve external stakeholders, regulators, and national coordination structures. This scalability is particularly relevant because cyber resilience requirements differ across sectors and maturity levels, yet each organization must demonstrate effective preparedness and continuous improvement.

While cyber exercises can approximate important decision-making, coordination, and communication dynamics of real incidents, they are not equivalent to real incidents in terms of uncertainty, stakes, external pressures, and real-world consequences. In both cases, actors must build situational awareness under uncertainty, prioritize actions under time pressure, coordinate across roles and organizational boundaries, communicate decisions, and manage resource constraints. The crucial distinction is that exercises remove real-world harm while preserving the decision and coordination pressures that reveal practical weaknesses. This makes exercises especially suitable for identifying systemic gaps ,such as unclear responsibilities, ineffective escalation thresholds, communication bottlenecks, or missing technical and organizational resources, before these gaps materialize during an actual crisis.

From a governance and assurance perspective, exercises are also described as a management instrument for validating and strengthening organizational readiness. ISO guidance highlights that exercises support the identification of gaps and help determine the effectiveness of response and recovery strategies[18] . In practice, this means that exercises can be used to test whether policies, plans, procedures, training concepts, technical tooling, and inter-organizational agreements are operationally fit for purpose under stress conditions. Exercises thereby provide evidence to support structured improvement: they create a traceable basis for after-action analysis, remediation planning, and re-testing, moving preparedness from an abstract goal to a measurable capability-building cycle.

Beyond assurance and capability validation, cyber exercises provide significant pedagogical value. The ETH cyber report[19] emphasizes that scenario-based simulations are a key element of “active learning.” Instead of merely memorizing procedures or theoretical concepts, participants apply their knowledge in interactive, time-sensitive situations using familiar techniques and tools. This shift from passive learning to applied performance is essential in cyber crisis contexts, where effective response depends not only on what people know, but on how they act, individually and collectively ,under pressure. Active learning through simulation strengthens the practical understanding of

[18] International Organization for Standardization. ISO 22398:2013 Societal security — Guidelines for exercises. 2013.

[19] Cyber Defense Project (CDP), Center for Security Studies (CSS), ETH Zürich. CSS Cyber Defense Report - Cybersecurity and Cyberdefense Exercises. 2018.

incident patterns and escalation dynamics, while also improving cross-functional collaboration, decision quality, and crisis communication.

Consequently, large-scale cyber simulations recreate cyber incidents in a controlled environment and enable participants to experience conditions that approximate real-world complexity without incurring real-world damage[20]. They provide a structured mechanism to generate learning and performance evidence, align stakeholders on roles and responsibilities, and improve both individual and organizational capabilities. In this dissertation, cyber exercises are therefore treated not as isolated events, but as recurring elements of a continuous preparedness cycle ,training, evaluation, and improvement ,that directly contributes to organizational and national cyber resilience by enabling faster, more coordinated, and more reliable crisis performance over time[21] .

Used terminology

This subchapter defines the key terms used throughout the dissertation. The purpose of this terminology section is not merely descriptive. It is to establish a consistent analytical vocabulary for the dissertation and to distinguish cyber exercises from adjacent forms of resilience validation, testing, and simulation. This is necessary because the literature and professional practice frequently use overlapping terms inconsistently, which weakens comparability and may lead to conceptual ambiguity in research and governance design.

For the purposes of this dissertation, the following terms are used in the sense defined below. See an overview in Figure 9.

Exercise:
An exercise is a structured and purposeful activity designed to train, assess, validate, and improve organizational performance under simulated or controlled conditions. Exercises may be used to verify plans, procedures, roles, communication paths, technical measures, decision-making routines, and inter-organizational coordination mechanisms. In this dissertation, the term exercise is used as a broad umbrella concept that includes different formats, such as discussion-based, functional, technical, and integrated exercises.

Cyber exercise:
A cyber exercise is defined in this dissertation as a structured, planned, and controlled activity in which realistic cyber incidents or cyberattack scenarios are simulated in order to train, assess, and improve the ability of an organization and its relevant stakeholders to detect, analyze, respond to, contain, mitigate, recover from, and learn from cyber incidents. A cyber exercise does not only rehearse technical response actions. It also tests organizational, managerial, legal, communicational, and inter-organizational dimensions of cyber incident and crisis management. In this sense, cyber exercises function both as preparedness instruments and as evaluation mechanisms. They help organizations validate roles and responsibilities, test escalation and coordination paths, improve playbooks and procedures, and identify weaknesses before a real incident occurs.

At the same time, cyber exercises should not be equated with real incidents. While they can approximate important decision-making, coordination, communication, and time-pressure dynamics of real cyber crises, they do not fully reproduce the uncertainty, consequences, external pressures, and stakes of the real incident environment. Their value lies precisely in creating a safe but demanding setting in which selected features of operational reality can be observed, practised, and evaluated.

[20] Cyber Defense Project (CDP), Center for Security Studies (CSS), ETH Zürich. CSS Cyber Defense Report - Cybersecurity and Cyberdefense Exercises. 2018.
[21] Finnish National Cyber Security Centre. Instructions for organising cyber exercises. s.l. : Traficom publications, 2020.

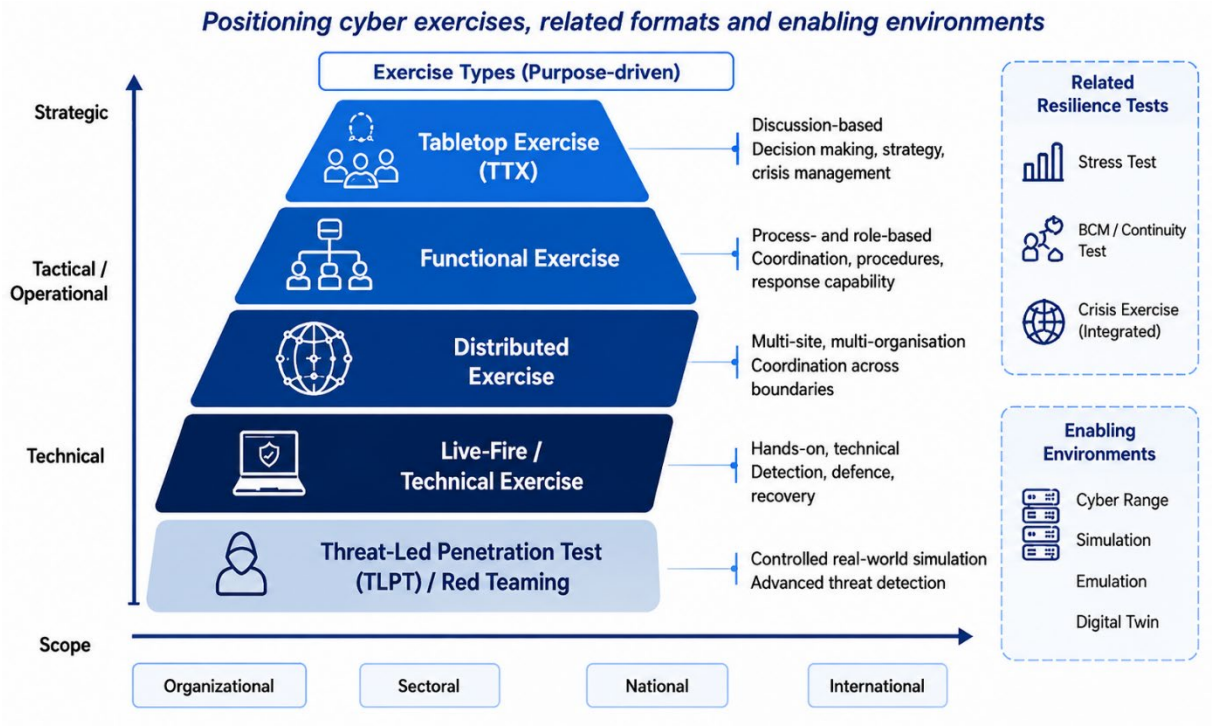


FIGURE 9 - CYBER EXERCISE AND RESILIENCE VALIDATION ECOSYSTEM

Stress test:
A stress test is a structured validation activity designed to examine how an organization, system, sector, or governance arrangement performs under adverse, high-pressure, or degraded conditions. In the cyber context, stress tests focus less on routine operational handling and more on resilience under strain, such as resource limitations, service degradation, dependency failure, regulatory pressure, or multiple concurrent disruptions. A stress test may overlap with a cyber exercise, but it is analytically distinct because its main purpose is to probe tolerance thresholds, breaking points, continuity capacity, and systemic interdependencies rather than primarily to train participants. In this dissertation, stress tests are treated as an adjacent but only partially overlapping category within the broader cyber resilience validation ecosystem.

Resilience testing / resilience validation:
Resilience testing, or resilience validation, is used here as a broader umbrella concept referring to structured activities through which the ability of an organization or system to withstand, respond to, recover from, and adapt to disruption is examined. In contrast to cyber exercises, which typically focus on participant action and organizational performance during a scenario, resilience testing may also include technical testing, continuity validation, dependency analysis, recovery verification, and intelligence-led security assessments. Cyber exercises are therefore treated in this dissertation as one important subset within the wider field of resilience validation.

Cyber range:
A cyber range is a controlled technical environment used to simulate networks, systems, infrastructures, or operational settings for training, experimentation, testing, and exercising in cybersecurity. Cyber ranges may support live technical drills, adversary emulation, hands-on defence training, malware analysis, or complex cyber exercises. However, a cyber range is not itself an exercise. It is an enabling environment or platform. In other words, a cyber exercise may be conducted on a cyber range, but the range and the exercise are conceptually distinct: the range provides the technical setting, while the exercise provides the objectives, scenario logic, participant roles, evaluation design, and governance framework.

Simulation, emulation, and digital twin:

These three terms are related but not interchangeable.

Simulation refers to the abstract or model-based representation of processes, systems, behaviours, or events. In cyber and resilience contexts, simulation is often used to reproduce selected aspects of operational reality without reproducing the full technical system itself. Simulations are useful when the goal is to model behaviour, dependencies, timing, or consequences at a manageable level of abstraction.

Emulation refers to the reproduction of the behaviour of a system or component in a way that mimics its functional characteristics more closely than a high-level simulation. In cyber environments, emulation is often used when the exercise or test requires more realistic technical interaction with services, devices, or protocols than a purely abstract simulation can provide.

A digital twin is a dynamic digital representation of a physical system, process, or environment that is linked to real or representative operational data and can be used for analysis, monitoring, scenario exploration, or testing. In resilience and exercise contexts, digital twins may support more realistic scenario development and environment modelling, especially where operational dependencies and system states matter over time.

In this dissertation, simulation, emulation, and digital twin are treated as different technical modelling approaches that may support cyber exercises or resilience testing, but they are not themselves equivalent to exercises.

Tabletop exercise (TTX):
A tabletop exercise is a discussion-based exercise format in which participants analyse and respond to a hypothetical cyber incident or crisis scenario through facilitated discussion. Tabletop exercises are primarily used to validate decision-making, escalation logic, communication, governance roles, and coordination processes rather than hands-on technical operations. They are especially relevant for executive, crisis-management, and cross-functional validation.

Functional exercise:
A functional exercise is a more operationally oriented exercise in which participants execute their actual roles and processes in a time-structured simulated event. The emphasis lies on process execution, coordination, workflow performance, and decision implementation rather than on debate alone. Functional exercises are particularly useful for validating incident response, continuity, crisis-management interaction, and inter-organizational cooperation under time pressure.

Red teaming:
Red teaming is a structured adversarial activity in which a designated team acts as an attacker in order to challenge defensive assumptions, expose weaknesses, and test response capability. In a broad sense, red teaming may occur within an exercise context, but it is not synonymous with cyber exercise as such. In this dissertation, red teaming is understood as an adversarial method that may form part of a technical or integrated exercise, especially where the purpose is to create realistic attacker pressure.

Adversary emulation:
Adversary emulation refers to the systematic reproduction of the tactics, techniques, and procedures of a specific threat actor or threat class. Compared to generic red teaming, adversary emulation is more intelligence-driven and behaviourally grounded. It aims to reproduce realistic attacker tradecraft rather than simply “behaving like an attacker” in an unspecific way. In this dissertation, adversary emulation is treated as a more specific and threat-informed adversarial method that may be embedded in advanced cyber exercises or resilience testing formats.

Threat-Led Penetration Testing (TLPT):
Threat-Led Penetration Testing is an intelligence-led, controlled, and production-relevant advanced testing approach in which realistic adversarial pressure is applied to critical functions and the systems supporting them. In the European regulatory context, TLPT is particularly associated with DORA and builds methodologically on the TIBER-EU approach. TLPT is not equivalent to a tabletop exercise, a generic penetration test, or a conventional cyber exercise. Rather, it is best understood as a specialised form of resilience-oriented adversarial testing that overlaps with cyber

exercise logic in certain respects, especially where detection, response, crisis coordination, third-party dependencies, and recovery are examined under realistic pressure. In this dissertation, TLPT is therefore treated as an adjacent but distinct category within the broader resilience validation ecosystem.

Scenario:
A scenario is a structured and plausible representation of a cyber incident or crisis situation that provides the narrative and operational frame for participant action. A scenario defines the threat context, affected systems or functions, actors involved, escalation dynamics, assumptions, and intended decision points that shape what participants are expected to perceive, decide, communicate, and execute. In this dissertation, the scenario is understood as the backbone that links exercise objectives, inject logic, participant roles, and evaluation criteria.

Inject:
An inject is a discrete piece of information, stimulus, event, or task introduced into an exercise in order to prompt participant action, decision-making, communication, or coordination. Injects are used to structure the progression of a scenario and to test whether intended objectives, reactions, and capability elements are triggered during the exercise. In more complex exercises, injects are often organised within a Master Scenario Event List (MSEL).

Master Scenario Event List (MSEL):
A Master Scenario Event List is a structured planning and control artefact used to organise the sequence of events, incident developments, and injects during an exercise. The MSEL helps planners and controllers manage timing, dependencies, escalation, and consistency across the exercise. In this dissertation, the MSEL is treated as one of the core artefacts of rigorous exercise preparation and controlled scenario delivery.

Player, planner, evaluator, and facilitator:

A player is a participant who reacts to exercise events in the role assigned to them and performs actions as they would in the real organizational context.
A planner is a person responsible for designing, organising, and preparing the exercise, including objectives, scope, scenario logic, artefacts, and coordination arrangements.
An evaluator is a person responsible for observing exercise performance, collecting evidence, and assessing the extent to which exercise objectives and expected capability elements were addressed.
A facilitator or controller is a person responsible for guiding the conduct of the exercise, delivering injects, maintaining flow, and ensuring that the scenario develops in a controlled and coherent manner.

Working distinction used in the dissertation.

Based on the above, this dissertation applies the following working distinction. Cyber exercises are treated as structured preparedness and evaluation activities focused on human, organizational, and inter-organizational performance under simulated cyber incident conditions. Stress tests and TLPT are treated as adjacent but analytically distinct forms of resilience validation, each with a stronger emphasis on pressure, thresholds, or adversarial realism. Cyber ranges, simulations, emulations, and digital twins are treated as enabling environments or technical modelling approaches that may support exercising and validation but are not themselves equivalent to exercises. This distinction is essential for the dissertation because governance design, comparability, lifecycle modelling, and AI integration all depend on conceptual clarity about what exactly is being planned, evaluated, and compared.

Analytical Dimensions for Comparing Cyber Exercises

The international cyber exercise landscape analysed in this chapter demonstrates that the term *cyber exercise* encompasses a wide range of activities that differ substantially in purpose, scale, governance, technical depth, participating communities, and evaluation approach. A strategic tabletop exercise involving government decision-makers, a technical live-fire exercise conducted in a cyber range, and a multinational operational exercise may all be described as cyber exercises, although they test different capabilities and generate fundamentally different forms of evidence. Direct comparison between such activities is therefore only meaningful if the characteristics that account for these differences are explicitly identified.

For this reason, the purpose of this subsection is not yet to establish a normative cyber exercise taxonomy. Instead, it develops a set of **analytical dimensions** that make the selected international and national reference exercises systematically comparable. These dimensions are derived inductively from the exercise formats examined in this chapter and from the terminology and distinctions introduced in Section 1.2.2. They provide the descriptive basis for the comparative analysis in Section 1.2.4 and subsequently inform the formal Cyber Exercise Taxonomy and Minimum Dataset developed at the national research layer in Section 2.3.

The first analytical dimension concerns **institutional ownership and governance**. Cyber exercises may be organised by individual organizations, sectoral bodies, national authorities, military organizations, EU institutions, international organizations, universities, specialised cyber range providers, or partnerships involving several of these actors. Institutional ownership is relevant because it influences exercise objectives, authority structures, participation models, information-sharing arrangements, and the extent to which findings can be translated into wider policy or capability-development processes.

A second dimension concerns the **scope or attribution level of the exercise**. For analytical purposes, exercises can be positioned at organizational, sectoral, national, or international level. Organizational exercises primarily test internal capabilities and interfaces. Sectoral exercises focus on shared dependencies, sector-specific risks, and coordination between entities providing related services. National exercises introduce broader questions of public-private coordination, government decision-making, national situational awareness, and cross-sector dependencies. International exercises additionally test interoperability, multinational information sharing, and coordination across national governance structures.

The third dimension concerns the **primary capability objective**. Exercises may focus on technical incident response, detection and containment, crisis-management coordination, executive decision-making, communication, continuity and recovery, legal or regulatory procedures, third-party dependencies, military mission objectives, or combinations of these areas. Explicitly recording the capability objective is essential because exercises should be compared according to what they are intended to validate rather than according to their label alone.

A fourth dimension concerns the **exercise format and execution model**. The international landscape includes discussion-based tabletop exercises, functional exercises, distributed exercises, live-fire or technical exercises, red-team/blue-team activities, threat-led penetration testing, full-scale simulations, and hybrid combinations. These formats differ considerably in preparation effort, realism, evidence-generation capacity, and the type of participant behaviour that can be observed. Consequently, exercise format must be treated as a primary analytical variable.

A fifth dimension concerns the **level of performance being assessed**. Some exercises primarily validate technical response, while others address tactical-operational coordination, strategic or executive decision-making, or several levels simultaneously. This dimension is particularly relevant

to large-scale and national exercises because meaningful cyber resilience increasingly depends on the interaction between technical, operational, and strategic layers.

A sixth dimension concerns the **sectoral and ecosystem context**. Exercises may focus on a single organization, a specific critical sector, several interdependent sectors, defence environments, public administration, or a broader whole-of-society ecosystem. Sectoral positioning is significant because threat scenarios, continuity requirements, regulatory obligations, and dependency structures differ substantially between areas such as energy, healthcare, finance, transport, public administration, and defence.

A seventh dimension concerns the **scenario and threat context**. Cyber exercise scenarios may involve ransomware, destructive attacks, supply-chain compromise, espionage, hybrid threats, operational technology disruption, state-sponsored activity, large-scale service failure, or multi-vector crises. Scenario complexity influences both the capabilities tested and the validity of exercise outcomes. Recording the threat context therefore helps distinguish between exercises that may otherwise appear similar based solely on format.

An eighth dimension concerns **scale and temporal characteristics**, including duration, number and type of participating organizations, recurrence, and exercise cadence. A recurring annual or multi-annual exercise series offers different possibilities for institutional learning than an isolated event. Recurrence is especially important for the dissertation because systematic improvement requires the possibility of revisiting previous findings and testing whether corrective measures have been implemented.

Finally, the analysis considers the **evaluation and improvement orientation** of the exercise. This includes whether objectives are explicitly evaluated, whether evidence is collected systematically, whether an After-Action Report or equivalent product is produced, whether findings are translated into corrective actions, and whether subsequent exercises provide an opportunity for re-testing. This dimension links the descriptive international analysis directly to the dissertation's broader concern with comparability and resilience-oriented improvement.

Taken together, these analytical dimensions provide a consistent lens for examining otherwise heterogeneous cyber exercise formats. They do not yet constitute the final taxonomy proposed by the dissertation. Their purpose in Part I is analytical: they make fragmentation visible, enable structured comparison of reference exercises, and identify the stable attributes required for a more formal comparability model. The constructive step is undertaken in Section 2.3, where these analytical observations are combined with the national governance requirements derived from the Austrian reference case and transformed into a formal **Cyber Exercise Taxonomy and Minimum Dataset**.

The purpose of this categorisation and database structure is not only descriptive. It also provides the empirical and conceptual basis for the taxonomy developed in Chapter 3.3 and for the comparability and measurement logic operationalized in the programme blueprint in Chapter 3.4.

ANALYTICAL DIMENSION	MAIN QUESTION
GOVERNANCE / ORGANISER	Who owns and governs the exercise?
SCOPE	Organizational, sectoral, national, or international?
CAPABILITY OBJECTIVE	What is intended to be tested?
EXERCISE FORMAT	How is the exercise conducted?
ASSESSED LEVEL	Technical, operational, strategic, or multi-level?
SECTOR / ECOSYSTEM	Which functions or dependencies are represented?
SCENARIO / THREAT	What type of disruption or adversary is simulated?

SCALE / CADENCE	How large and how recurrent is the exercise?
EVALUATION / IMPROVEMENT	What evidence, findings, and follow-up are generated?

TABLE 1 - ANALYTICAL DIMENSIONS USED FOR THE COMPARATIVE ASSESSMENT OF CXS

Comparative Mapping of Selected Cyber Exercise Refence Cases

The analytical dimensions established in Section 1.2.3 are applied in this subsection to a selected set of recurring cyber exercise and resilience-validation initiatives. The cases were selected because they represent different institutional levels, exercise formats, capability objectives, technical depths, governance models, and geographical scopes. Together, they provide a sufficiently diverse basis for examining both the fragmentation of contemporary cyber exercise practice and the recurring characteristics that may support greater comparability.

The purpose of the comparison is not to rank the selected exercises or to identify a single best-practice format. Cyber exercise effectiveness depends strongly on context, including purpose, maturity, available resources, participating organizations, technical environment, confidentiality requirements, and the capability layer being tested. A strategic tabletop exercise designed to test executive decision-making cannot be assessed according to the same criteria as a live-fire technical exercise designed to test detection, containment, and operational defence under adversarial pressure. Similarly, a national crisis-management exercise and a multinational military cyber exercise may share certain governance characteristics while serving substantially different operational objectives.

The comparison therefore focuses on the structural characteristics that make heterogeneous exercises analytically distinguishable. These include institutional ownership and governance, primary capability objectives, exercise format, assessed capability level, sectoral or ecosystem context, geographical and governance scope, recurrence, and the orientation towards evaluation and improvement. Applying the same analytical dimensions across different exercise families makes it possible to identify where meaningful differences remain and where common principles begin to emerge.

The selected cases include European-level crisis-management exercises, multinational and military exercise programmes, national crisis-management and preparedness exercises, technically intensive cyber range-supported exercises, and advanced resilience-testing formats. These comprise ENISA Cyber Europe and Blue OLEx; NATO and NATO CCDCOE exercises such as Cyber Coalition, Locked Shields, and Crossed Swords; the TIBER-EU framework as an adjacent resilience-validation model; the U.S. CISA Cyber Storm exercise series; USCYBERCOM exercises such as Cyber Flag and Cyber Guard; Germany's LÜKEX; Finland's KYHA national cyber exercises; and the Austrian KSÖ Cybersecurity Planspiele. The current dissertation already analyses these exercise families as representative reference cases of the international cyber exercise landscape (see Overview in Figure 10).

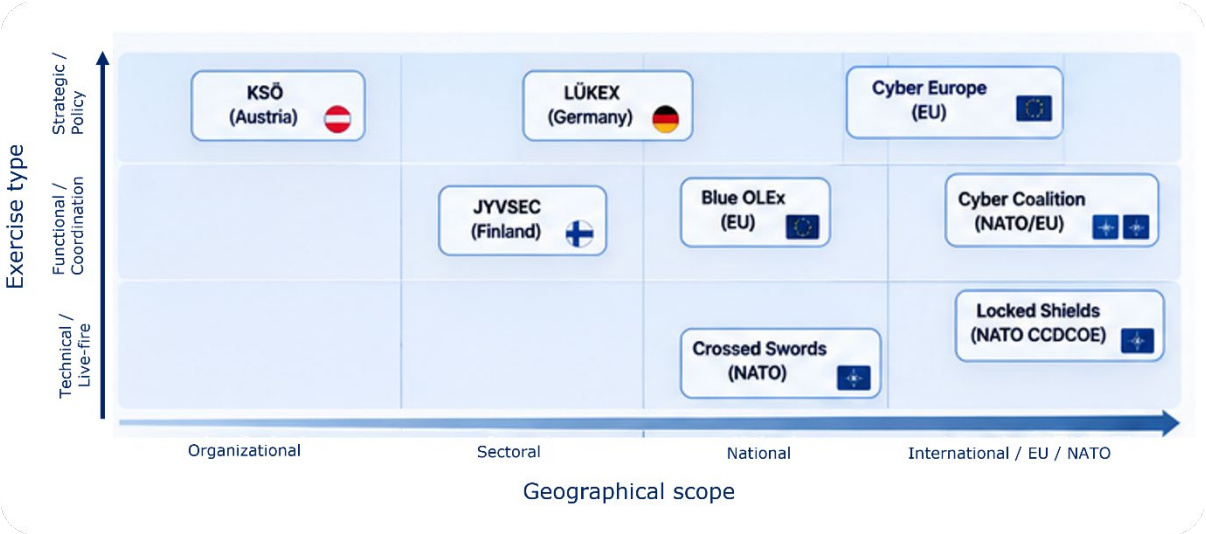


FIGURE 10 - POSITIONING SELECTED CXS BY SCOPE AND EXERCISE TYPE

ENISA Cyber Europe represents a recurring European large-scale cyber crisis-management exercise focused primarily on cross-border coordination, information sharing, common situational awareness, sectoral resilience, and decision-making across Member States and EU institutions. Its significance for this dissertation lies less in its technical intensity than in its ability to exercise the interfaces between national and European actors during large-scale cyber crises. The exercise therefore illustrates the growing importance of operational and strategic coordination as a component of cyber resilience.

Blue OLEx provides a complementary European reference point. Its focus is more explicitly executive and operational, particularly on cooperation between national cybersecurity crisis-management authorities and EU-level structures such as EU-CyCLONe. Compared with technically intensive cyber exercises, Blue OLEx demonstrates the importance of exercising escalation, shared situational awareness, information exchange, and executive decision-making. It therefore represents a governance-intensive exercise format in which the primary object of validation is not individual technical performance but the ability of authorities to coordinate effectively under crisis conditions.

NATO Cyber Coalition provides a multinational defence-oriented reference case focused on collective cyber defence, interoperability, coordination, and the ability of NATO structures, Allies, and partners to operate together during significant cyber incidents. The exercise illustrates the requirements associated with multinational command-and-control, information exchange, procedural interoperability, and coordinated response across distributed participants. Its recurring nature also makes it relevant as an example of institutionalised exercise-based capability development rather than a one-off training activity.

NATO CCDCOE Locked Shields represents a substantially more technically intensive exercise family. It combines live-fire defensive activity with operational, legal, strategic, and communication dimensions. Blue Teams defend complex simulated national infrastructures against sustained adversarial activity while simultaneously managing broader crisis implications. The relevance of Locked Shields lies in its integration of multiple capability layers within the same exercise environment. Technical evidence, operational coordination, legal decision-making, and strategic communication are exercised in parallel, demonstrating how technically intensive exercises may also generate governance-relevant findings.

Crossed Swords provides a further military reference case but differs from Locked Shields in purpose. It focuses more strongly on the planning and execution of cyber operations in a crisis or conflict environment and incorporates offensive and defensive dimensions. The exercise is

particularly relevant because it demonstrates the integration of cyber operations into wider military planning, legal constraints, mission objectives, and multi-domain coordination. It therefore illustrates how exercise design is shaped by the operational purpose of the exercise and why a single universal exercise classification would be insufficient.

TIBER-EU is included in the comparative landscape as an adjacent resilience-validation mechanism rather than as a conventional cyber exercise. Its relevance lies in its structured, threat-intelligence-led approach to testing critical functions and live production environments under controlled adversarial conditions. TIBER-EU demonstrates a high level of formal governance, predefined roles, controlled execution, structured reporting, remediation, and follow-up. It is therefore useful for this dissertation as evidence that resilience-testing regimes are increasingly oriented towards traceability, evidence, accountability, and verified improvement, even where their methodological form differs from traditional exercise formats.

The U.S. CISA Cyber Storm series represents a national-scale functional and crisis-management exercise model. It focuses on coordination between government, critical infrastructure operators, and other stakeholders during major cyber incidents. Cyber Storm is particularly relevant to the national governance discussion because it exercises the relationships between organizations rather than focusing exclusively on technical defensive activity. Its structured after-action reporting further illustrates how large-scale exercises can contribute to institutional learning across recurring programme cycles.

USCYBERCOM's Cyber Flag and Cyber Guard series provide additional defence-oriented reference cases. Cyber Flag emphasises operational readiness, multinational interoperability, and cyber mission performance, while Cyber Guard places greater emphasis on command-level integration and broader operational coordination. Together, these exercises illustrate how a defence organization may maintain a portfolio of different exercise types aligned to different readiness and capability objectives rather than relying on a single exercise format.

Germany's LÜKEX represents a national strategic crisis-management exercise model. Unlike technically focused cyber range exercises, LÜKEX primarily tests whole-of-government crisis-management structures, political-administrative coordination, communication, and cooperation between federal and state actors. Cyber-related scenarios such as LÜKEX 23 demonstrate that national cyber resilience also depends on the state's ability to maintain governance and decision-making under disruption. This makes LÜKEX particularly relevant for understanding cyber exercises as instruments of state-level resilience rather than only cybersecurity training.

Finland's KYHA exercises provide a different national model. They combine recurring national exercise delivery with technically realistic cyber range infrastructure and multi-entity participation. The Finnish model is particularly relevant because it integrates public-sector preparedness, critical infrastructure participation, technical-operational exercising, research, education, and dedicated exercise infrastructure. The recurring nature of the KYHA portfolio and the role of JYVSECTEC demonstrate how national cyber exercising can be institutionalised as a sustained capability rather than organised only in response to individual events.

The Austrian KSÖ Cybersecurity Planspiele provide an additional national reference case and are especially relevant because Austria serves as the principal reference environment in Part II of the dissertation. The KSÖ exercises bring together public authorities, critical infrastructure operators, private-sector participants, research organizations, and technical partners. Across different iterations, they have incorporated crisis-management, cross-organizational coordination, hybrid-threat scenarios, and cyber range-supported elements. Their inclusion in Part I therefore creates a direct bridge between the international comparison and the deeper Austrian governance analysis undertaken in Part II.

Exercise	Governance / Organiser	Primary capability focus	Exercise format	Assessed level	Scope	Cadence
ENISA Cyber Europe	ENISA with EU Member States and EU bodies	Cross-border crisis coordination, information sharing, sector resilience	Distributed / hybrid crisis-management exercise	Operational + Strategic	EU / International	Biennial
ENISA Blue OLEx	ENISA, EU-CyCLONe, Member State cyber crisis authorities	Executive coordination, escalation, shared situational awareness	Executive tabletop	Operational + Strategic	EU / International	Recurring
NATO Cyber Coalition	NATO Allied Command Transformation	Coalition interoperability, collective cyber defence, coordination	Distributed operations-based exercise	Operational + Strategic	International / Alliance	Annual
Locked Shields	NATO CCDCOE	Defence of complex infrastructure under active adversarial pressure	Live-fire technical-operational exercise	Technical + Operational + Strategic	International	Annual
Crossed Swords	NATO CCDCOE	Full-spectrum cyber operations and military cyber coordination	Operations-based cyber exercise	Tactical + Operational	International / Defence	Annual
CISA Cyber Storm	U.S. CISA	National cyber crisis coordination and critical-infrastructure response	Distributed functional exercise	Operational + Strategic	National with international participation	Biennial
LÜKEX	German Federal Government, Länder and BBK	Whole-of-government crisis management and cross-sector coordination	Strategic staff / crisis-management exercise	Strategic + Operational	National	Periodic
KYHA	JYVSECTEC/JAMK with Finnish national support	National preparedness, technical-operational response and multi-entity coordination	Range-enabled technical-operational exercise	Technical + Operational + Leadership	National	Multiple exercises annually
Cyber Flag	USCYBERCOM	Multinational cyber operations, interoperability and mission readiness	Operations-based field training	Tactical + Operational	International / Defence	Recurring
Cyber Guard	USCYBERCOM and Joint Staff	Command-level cyber readiness and operational integration	Operations-based command exercise	Operational + Strategic	National / Defence	Annual

TABLE 2 - COMPARATIVE MAPPING OF SELECTED CX REFERENCE CASES

The comparative mapping demonstrates that contemporary cyber exercises cannot be meaningfully differentiated through a single typological dimension. Similar labels may refer to exercises with fundamentally different governance structures, capability objectives, execution formats, technical depth, and geographical scope. Conversely, exercises originating from different institutional traditions often exhibit recurring structural characteristics, particularly regarding clearly defined objectives, multi-level coordination, structured scenario control, evaluation, and recurring exercise cycles. The comparison therefore confirms both components of the analytical problem addressed by H1: substantial fragmentation remains at the level of exercise form and implementation, while common principles increasingly emerge at the level of methodology, governance, evidence, and improvement.

These observations provide the empirical basis for the analytical dimensions used throughout Part I. They do not yet constitute the formal Cyber Exercise Taxonomy proposed by the dissertation. The latter is developed in Section 2.3, where the international evidence is combined with the European convergence findings and the Austrian reference case to construct a taxonomy and minimum dataset suitable for national cyber exercise governance.

Across the selected cases, several recurring characteristics emerge. More mature exercise programmes tend to define institutional ownership clearly, distinguish between strategic, operational, and technical objectives, use recurring exercise cycles rather than isolated events, rely on structured scenario and control mechanisms, and incorporate some form of evaluation and post-exercise learning. Large national and international exercises additionally place strong emphasis on inter-organizational coordination, dependency management, information sharing, and common situational awareness. Technically intensive formats introduce additional requirements concerning controlled execution environments, adversarial realism, safety, instrumentation, and observable technical evidence.

The analysis therefore reveals an important distinction between standardisation of exercise form and standardisation of exercise description. Standardising exercise form would be neither realistic nor desirable because different capability objectives require different formats, levels of realism, participants, and technical environments. An executive crisis-management tabletop exercise should remain fundamentally different from a live-fire technical defence exercise if the two are intended to validate different capabilities. Standardisation of exercise description, however, is both possible and necessary if national stakeholders wish to understand exercise portfolios, compare activities across organizations and sectors, identify capability gaps, and track development over time.

This distinction provides a direct analytical bridge to the national research layer. Part II does not seek to standardise cyber exercises themselves. Instead, Section 2.3 develops a Cyber Exercise Taxonomy and Minimum Dataset that standardise the information used to describe, compare, and govern them. In this way, legitimate diversity in exercise design can be preserved while comparability, portfolio visibility, evidence aggregation, and longitudinal improvement tracking are strengthened at national programme level.

The findings of this comparative analysis also contribute directly to the first research gap identified in the dissertation. The problem is not simply that many different types of cyber exercise exist; such diversity is necessary and reflects legitimate differences in purpose and context. The comparability gap arises because these differences are not always documented through sufficiently consistent descriptors, artefacts, and evaluation structures. Consequently, exercise results cannot automatically be aggregated or interpreted as comparable evidence of resilience. The analytical dimensions developed in this chapter therefore provide the first step towards addressing that gap without imposing unnecessary uniformity on exercise design.

1.2.1.1 *ENISA Cyber Europe*

Cyber Europe is ENISA's biennial series of large-scale, cross-border cyber crisis management exercises, organised since 2010. The exercises are designed to simulate major cybersecurity incidents that escalate into cyber crises, using complex scenarios inspired by real events and threats. Cyber Europe is developed and delivered in collaboration with European cybersecurity experts and planners from participating countries and EU bodies, with the explicit aim of strengthening Europe's collective preparedness and resilience through realistic coordination under pressure.

From an exercise-typology perspective, Cyber Europe is primarily a crisis management and coordination exercise at European scale. It is not a single "live-fire" technical range event focused only on hands-on defence; instead, it tests how national and EU-level stakeholders share information, build a common operational picture, coordinate decisions, and manage cascading impacts across multiple organizations and jurisdictions. This makes it particularly relevant for dissertation work on national and multinational exercise ecosystems, because it represents a mature reference model for exercising governance, interoperability, and cross-border cooperation rather than isolated organizational response.

A characteristic design feature of the series is that each edition typically emphasises a sectoral or ecosystem context to surface real dependencies. For example, Cyber Europe 2022 centred on the healthcare ecosystem, bringing together stakeholders such as national CSIRTs, cybersecurity authorities, ministries, healthcare organizations, and relevant service providers to test the resilience and coordination mechanisms needed during a complex cyber crisis. The Cyber Europe 2024 edition focused on the energy sector and was conducted in June 2024, explicitly aiming to identify gaps and strengthen cybersecurity preparedness and resilience in a high-impact, geopolitically relevant context.

In terms of lifecycle and governance, Cyber Europe reflects an end-to-end methodology consistent with ENISA's broader exercise doctrine: it involves a substantial planning and design effort with multinational planners, definition of objectives and scope, scenario development and controlled inject delivery during execution, and systematic evaluation. ENISA consolidates results into an After-Action Report (AAR) that documents the exercise set-up, methodology, and the key findings and improvement themes derived from participant feedback and evaluation activities. The Cyber Europe 2024 AAR, for instance, positions the exercise explicitly as a mechanism to identify gaps and improve preparedness at European scale, and it describes structured data collection and evaluation as part of the exercise cycle.

For cyber exercise research and practice, Cyber Europe demonstrates several transferable lessons. First, it shows how large-scale exercises can operationalise cross-border coordination by bringing together the technical layer (e.g., incident response communities) with operational and strategic decision layers, thereby testing the translation of technical incident realities into actionable crisis management. Second, it provides a reference for sector-focused scenario design: choosing a critical ecosystem (health, energy) exposes dependencies, supply-chain issues, and cascading effects that are difficult to observe through single-organization exercises. Third, the series reinforces a continuous-improvement logic: the production of AARs and the emphasis on identifying gaps and strengthening preparedness create an evidence base for iterative refinement of procedures, cooperation mechanisms, and future exercise design.

1.2.1.2 *ENISA Blue OLEx*

Blue OLEx (Blueprint Operational Level Exercise) is ENISA's recurring, executive-level tabletop exercise series designed to test and strengthen how the EU coordinates during large-scale, cross-border cybersecurity incidents and crises. The series started in 2019 as a "first-of-its-kind" exercise focusing on the operational layer of the EU cyber crisis response framework (the "Blueprint"), bringing together the heads of national cybersecurity authorities from EU Member States, the European Commission, and ENISA. Over subsequent editions, Blue OLEx evolved into the principal executive exercising venue for the European cyber crisis liaison organization network (EU-CyCLONe), the operational-level cooperation network for national authorities in charge of cyber crisis management.

Blue OLEx is best understood as an exercise that targets the operational and strategic decision layer rather than hands-on technical defence. Its primary objective is to assess and improve cooperation among crisis management executives and directors across Member States and EU institutions, with particular emphasis on the mechanisms that must work under time pressure: information sharing, creation of shared situational awareness, escalation criteria, and coordinated decision-making. ENISA explicitly links the exercise outcomes to the continuous development of the EU's crisis management arrangements: lessons learned from Blue OLEx have been used to refine the standard operating procedures (SOPs) of EU-CyCLONe and to inform broader policy work on EU-level cyber crisis management.
The governance relevance of Blue OLEx is strengthened by its legal and institutional embedding after NIS2. Under Directive (EU) 2022/2555 (NIS2), EU-CyCLONe is established to support the coordinated management of large-scale incidents and crises at operational level, and ENISA is mandated to provide the secretariat, support secure information exchange, and provide tools for cooperation. In this setting, Blue OLEx functions as a practical method to "exercise the system": it provides a controlled environment in which senior officials can rehearse cross-border coordination routines, clarify thresholds for escalation, and validate whether cooperation mechanisms can produce a coherent operational picture suitable for political-level decision support.
Recent iterations illustrate the exercise's focus areas and its connection to EU crisis management development. The 2024 edition convened high-level executives from Member States' competent authorities responsible for cyber crisis management and/or cyber policy, the European Commission, and ENISA, with the CSIRTs Network liaison coordinator involved as the representative of the technical layer; strategic discussions included effective information sharing and criteria for escalation, and the lessons learned were explicitly positioned as input to the then ongoing revision of the EU cyber crisis management Blueprint. The 2025 edition was explicitly framed as being driven by the revised Cyber Blueprint adopted earlier that year and focused on the role of executives in shaping coherent EU-level crisis management, again convening the same core community (Member States, Commission, ENISA) to test EU cooperation under simulated crisis pressure.

Blue OLEx is a useful reference case because it demonstrates how cyber exercising can be institutionalised as a governance mechanism: it is recurring, anchored in a formal EU cooperation network, and oriented toward producing actionable lessons for procedures and coordination. It also illustrates a mature approach to the multi-layer challenge of cyber crises, where technical incident handling must be translated into operational situational awareness and policy-relevant decision support. In this sense, Blue OLEx complements technically intensive live-fire exercises by focusing on the interoperability and coordination issues that determine whether Europe can act coherently during major cross-border cyber crises.

1.2.1.3 *Tiber EU*

TIBER-EU (Threat Intelligence-Based Ethical Red Teaming) is a European framework, initiated by the European Central Bank, for conducting **controlled, bespoke, intelligence-led red team tests** against an entity's **critical live production systems**. Its purpose is not "penetration testing for its own sake", but the structured validation and improvement of an organization's **protection, detection, and response** capabilities under realistic adversary behaviour, while maintaining strict safety and governance controls. TIBER-EU is designed as a common European approach that can be implemented by national authorities (via national "TIBER" programmes) and used to provide assurance to supervisors about the cyber resilience of entities under their responsibility.[22]
Methodologically, TIBER-EU resembles a high-fidelity **operations-based cyber exercise**: it simulates a real attacker campaign (end-to-end) against defined critical functions and systems, while the organization's defensive teams operate as in real life. A defining feature is that TIBER-EU tests are **threat-intelligence driven**. The test is based on an up-to-date, tailored threat intelligence assessment that identifies plausible adversaries, tactics, and attack paths relevant to the target organization and its ecosystem. This approach increases realism and ensures that the test measures capabilities against threats that matter, rather than generic checklists.
The TIBER-EU lifecycle is structured into **three mandatory phases**: **preparation, testing, and closure**. In the **preparation phase**, the test is formally launched, governance is set up, and the **scope is defined and attested by the entity's board**, then validated by the relevant authority. This phase also includes procurement/engagement of the **Threat Intelligence (TI) provider** and the **Red Team (RT) provider**, and the establishment of key internal roles. A core governance mechanism is the separation of teams: a **White Team** within the entity manages the test (scope, safety, internal coordination) and liaises with the authority's **TIBER Cyber Team (TCT)**, while the **Blue Team** (defenders) remains unaware during preparation and execution to preserve realism.
During the **testing phase**, the TI provider produces the threat intelligence deliverable(s) that inform realistic attack scenarios, and the RT provider executes the campaign within agreed rules, safety constraints, and scope. The RT provider plans and conducts the test and develops a **Red Team Test Plan** that is agreed with the entity's White Team and approved by the TCT, ensuring both realism and control. The Blue Team defends business-as-usual, enabling the test to assess detection, escalation, containment, and operational response in conditions that mimic a real intrusion.

The **closure phase** turns the test into measurable improvement. It includes structured reporting, reflection, and ,critically ,**remediation planning and results sharing**, so that findings are translated into concrete corrective actions. Many TIBER implementations also include a "replay" or learning component in which relevant defenders are later briefed on the attack chain to ensure that lessons are internalised and that improvements can be verified. In this sense, TIBER-EU institutionalises an improvement loop comparable to mature exercise programmes: **test → evidence → remediation → validation**.

TIBER-EU is particularly valuable because it provides a well-governed model for **high-realism cyber exercising with strong auditability**. It defines roles, deliverables, and approvals that reduce ambiguity (White Team governance, TCT oversight, TI and RT provider responsibilities) and it aligns technical testing with organizational learning and management accountability. It is also increasingly relevant in the EU regulatory environment because DORA's **Threat-Led Penetration Testing (TLPT)** expectations are widely seen as methodologically aligned with TIBER-EU and

[22] European Central Bank. (n.d.). TIBER-EU framework. https://www.ecb.europa.eu/paym/cyber-resilience/tiber-eu/html/index.en.html

national TIBER frameworks, making TIBER-EU a practical reference point for "advanced, intelligence-led" resilience exercises in the financial sector and beyond.[23]

1.2.1.4 *NATO Crossed Swords*

Crossed Swords (XS) is a NATO-associated cyber exercise conducted by the **NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE)** in Tallinn. Unlike large-scale defence-focused exercises such as Locked Shields, Crossed Swords is primarily oriented toward **cyberspace operations**, with a strong emphasis on training participants to plan and execute **full-spectrum operations** in a simulated crisis and conflict environment. CCDCOE describes Crossed Swords as a comprehensive exercise that aims to enhance the skills of **operational-level military command elements** in managing offensive and defensive cyberspace capabilities and of **tactical-level cyber specialists** in executing both defensive activities and full-spectrum offensive cyber operations.[24]

From an exercise typology perspective, Crossed Swords can be characterised as a **technical-operational, operations-based "live-fire" cyber operations exercise**. The exercise design places participants into roles that mirror military cyber operations planning and execution under realistic constraints (time pressure, contested environments, and mission objectives). NATO's Allied Command Transformation has highlighted that Crossed Swords is used to train national red teamers and cyber command specialists, and that its focus is on practising a range of offensive cyber operations in a context that is increasingly relevant for modern military operations.
Crossed Swords is also notable for integrating non-technical dimensions that are essential for realistic cyber operations. CCDCOE notes that beyond cyber warfare training, the exercise promotes broader multi-domain competencies such as **information operations**, **legal perspectives**, and cooperation with **Special Operations Forces**, reflecting the reality that operational cyber effects must be coordinated with legal oversight, strategic objectives, and communication considerations.[25]

Crossed Swords is a useful reference case for cyber exercise programmes because it demonstrates how an exercise can be designed around **mission-oriented cyber operations** rather than only defensive incident response. It illustrates (i) how to train command-and-control and planning functions alongside hands-on cyber activity, (ii) how to embed cyber activity into a wider crisis/conflict narrative, and (iii) how to combine technical realism with governance constraints (especially legal and coordination requirements). As such, Crossed Swords complements defence-oriented "infrastructure defence" exercises by focusing on the operational planning and execution layer that becomes relevant when cyber is treated as an integrated domain of military operations.

[23] European Central Bank. (2025, November). TIBER-EU SSM implementation guide: How to implement the TIBER-EU framework for the DORA TLPT of significant institutions. https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.supervisory_guide202511.en.pdf
[24] NATO Cooperative Cyber Defence Centre of Excellence. (n.d.). *Crossed Swords*. https://ccdcoe.org/crossed-swords/
[25] NATO Cooperative Cyber Defence Centre of Excellence. (n.d.). The NATO Cooperative Cyber Defence Centre of Excellence. https://ccdcoe.org/

1.2.1.5 *NATO Cyber Coalition*

Cyber Coalition[26] is NATO's flagship annual **collective cyber defence exercise**, planned and conducted by **Allied Command Transformation (ACT)**[27] under the governance of NATO's **Military Committee**. It is widely described as one of the world's largest cyber defence exercises and serves as a structured platform for Allies (and selected partners) to practise operating together in and through cyberspace, with a strong focus on interoperability, coordination, and the testing of procedures and capabilities in a controlled exercise environment.
From an exercise-typology perspective, Cyber Coalition is best characterised as an **operations-based, distributed cyber defence exercise** rather than a purely discussion-based tabletop. Participants typically train through realistic, multi-day scenario play where teams operate from their home organizations and designated exercise locations, using established incident response and operational coordination structures, while exercise control injects scenario developments and collects observations. Public reporting on Cyber Coalition 2024, for example, describes the event as run on a simulated network environment and executed in a distributed manner across participating teams, enabling "safe" experimentation and testing without the risks of a real-world incident.

The exercise's primary purpose is to strengthen NATO's collective cyber posture by training and testing **shared situational awareness, information-sharing mechanisms, incident response coordination, and the ability to conduct cyber operations together**. ACT frames Cyber Coalition as a venue to test and exchange on NATO cyber defence, to refine tactics, techniques and procedures, and to modernise capabilities through experimentation ,reflecting a broader NATO logic in which exercising is not only training, but also a driver for capability development and interoperability across the Alliance. In this sense, Cyber Coalition complements high-fidelity "network defence" live-fire formats (e.g., Locked Shields) by placing particular emphasis on **coalition coordination and operational integration**, including how cyber activities connect to wider multi-domain operations and command-and-control processes.

A distinctive feature of Cyber Coalition is its **broad participation model**, which typically includes NATO structures, Allied and partner cyber units, and (depending on the year and scenario) contributions from additional communities such as industry and academia. Public communications around Cyber Coalition 2025 emphasise the scale of participation (with reporting of more than a thousand cyber defenders involved) and note CCDCOE's support to the exercise, including storylines that incorporate operational/technical and legal dimensions. This illustrates that Cyber Coalition is not limited to technical incident handling alone, but also tests governance constraints and decision interfaces that matter in coalition operations.

Although public sources describe Cyber Coalition primarily through its execution week, the exercise implicitly follows a full lifecycle consistent with established exercise methodologies. A substantial **design and preparation period** precedes execution, during which ACT develops objectives, scenario architecture, control structures, and technical environments. The **execution phase** then operationalises these design choices through controlled inject delivery and distributed play. Finally, Cyber Coalition is explicitly positioned as a mechanism for capturing lessons and validating emerging concepts; participating organizations commonly use observations and outputs to refine internal procedures and contribute to Alliance-level learning about interoperability, information exchange, and coordinated cyber defence.

[26] NATO Allied Command Transformation. (n.d.). Cyber Coalition. https://www.act.nato.int/activities/cyber-coalition/
[27] NATO Allied Command Transformation. (n.d.). *Our work*. https://www.act.nato.int/our-work/

For dissertation purposes, Cyber Coalition is therefore a strong reference case for a **multinational, governance-relevant cyber exercise** that aims to improve cyber resilience at ecosystem scale. It demonstrates how an annual programme can institutionalise (i) **coalition-wide coordination routines**, (ii) repeatable testing of information-sharing and incident response mechanisms, and (iii) capability modernisation through structured experimentation. It also provides a practical example of how large-scale exercises can integrate technical and non-technical dimensions (including legal and operational constraints) while remaining feasible through distributed participation

1.2.1.6 *NATO CCDCOE Locked Shields*

Locked Shields[28] is NATO's flagship **live-fire** cyber defence exercise organised annually by the **NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE)**[29] **and CR14**[30]. It is widely regarded as one of the most complex, large-scale real-time cyber defence exercises, bringing together multinational **Blue Teams** formed by CCDCOE member nations and partners to defend a simulated national environment against an active **Red Team** adversary. Participants typically play the role of a national cyber rapid reaction capability supporting a fictional country during a major cyber crisis, with the exercise environment representing critical services and interconnected systems under sustained attack.

From an exercise-typology perspective, Locked Shields is best described as an **operations-based, technical-operational "network defence" exercise** with a strong **strategic layer**. While the core training audience is the defensive Blue Team, the exercise goes beyond technical incident handling and includes decision-making under pressure across multiple dimensions ,such as crisis leadership, legal considerations, and strategic communications ,reflecting the reality that large-scale cyber incidents are not only technical events but also governance and societal challenges.
A distinctive feature of Locked Shields is its multi-team structure and orchestration model. In addition to the **Red Team** (attackers) and **Blue Teams** (defenders), the exercise is supported by neutral enabling teams that build and operate the scenario and ensure fairness and continuity of play. Public descriptions commonly reference "control and enablement" teams (e.g., teams responsible for scenario development, simulation of normal network usage, and exercise governance/refereeing), which operationalises the separation between training audience and exercise control that is typical in high-fidelity exercises.

Although Locked Shields is executed over an intensive live-fire period, it implicitly follows a full exercise lifecycle consistent with established cyber exercise methodologies. First, there is a substantial **design and preparation phase**, during which CCDCOE engineers and maintains a complex simulated infrastructure and prepares attack and inject content aligned with exercise objectives. Second, during **execution**, Blue Teams defend availability and integrity of services, perform detection and response, and coordinate internal crisis processes while being scored on performance. Third, Locked Shields places emphasis on **structured evaluation and lessons learned**, reflected in the production of after-action reporting and the long-running iterative improvement of the exercise format across years.

Locked Shields is a useful reference model for national and multinational cyber exercise programmes because it demonstrates how a single exercise can integrate (i) **hands-on technical defence under realistic adversary behaviour**, (ii) **cross-functional crisis management**

[28] NATO Locked Shields: Locked Shields
[29] NATO CCDCOE: CCDCOE - The NATO Cooperative Cyber Defence Centre of Excellence is a multinational and interdisciplinary hub of cyber defence expertise.
[30] CR14: CR14 Homepage

(legal/communications/leadership decisions), and (iii) an **improvement-oriented cycle** supported by systematic evaluation. It therefore illustrates a mature approach to building collective cyber defence capability through repeated, high-fidelity exercising that tests not only tools and technical procedures but also coordination, decision authority, and resilience governance at scale. [31]

1.2.1.7 *U.S. USCYBERCOM*

United States Cyber Command (USCYBERCOM)[32] is the U.S. Department of Defence combatant command responsible for directing and coordinating cyberspace planning and operations to defend and advance U.S. national interests in collaboration with domestic and international partners. Its publicly stated mission includes defending Department of Defence information systems, supporting joint force commanders with cyberspace operations, and defending the nation from significant cyberattacks. This operational mandate provides the context for USCYBERCOM's major exercise programme, which is designed to develop and certify cyber mission readiness and to strengthen multinational and interagency coordination under realistic cyber threat conditions.

A central exercise series associated with USCYBERCOM is **Cyber Flag**, described by USCYBERCOM as its **biannual multinational cyber field training exercise**. Cyber Flag is explicitly designed to strengthen partnerships and improve the ability of participating nations to **detect, respond to, and defend** against malicious cyber activity. In other words, Cyber Flag functions as a coalition-focused operational exercise that emphasises interoperability and joint defensive performance rather than being limited to individual organizational training. USCYBERCOM has also used Cyber Flag to provide certification opportunities for mission teams; for example, official reporting on Cyber Flag 24-2 states that the exercise included a certification opportunity for **offensive mission teams** conducting offensive cyberspace operations and aimed to improve readiness and coordination among close partners. This illustrates that the series can address both defensive and offensive mission readiness depending on the edition and objectives, while remaining structured as a controlled training environment rather than real-world operations.

Complementing Cyber Flag, USCYBERCOM runs **Cyber Guard**[33], presented as a capstone-style exercise that tests higher-level integration and readiness in globally relevant scenarios. USCYBERCOM reporting describes Cyber Guard as simulating real-world cyber threats and being designed to sharpen unified defence and mission integration across participating entities. In this sense, Cyber Guard can be characterised as an exercise focused on **command-and-control alignment, operational coordination, and decision-making**, linking cyber operations to broader defence planning and multi-command collaboration. Public communication around Cyber Guard also highlights its role in improving collaboration across partners and aligning operational responses under complex, multi-dilemma conditions.

From a dissertation perspective, USCYBERCOM's exercise portfolio is valuable as a reference model for **large-scale, recurring, operations-based cyber exercises** that support collective defence and multinational readiness. The key contribution is not only technical training, but the systematic exercising of coalition interoperability: how multiple nations and organizations share situational awareness, synchronise defensive actions, and coordinate operational decisions in a realistic, time-pressured cyber incident environment. Cyber Flag demonstrates how such an exercise can be used as a recurring platform for improving joint defensive capability (and, in some

[31] An automated AI evaluations tool would be helpful to compare the participating teams. The usage of automated or even AI tools were allowed and used for detection and hardening as well. Here we can also see the increasing implementation of AI and automation usage for CXs. Source: I was participating the NATO Locked Shields Partner Run this year.

[32] U.S. Cyber Command: Home USCYBERCOM

[33] Cyber Guard 2025: U.S. Cyber Command Wraps Up Largest-Ever Cyber Guard Exercise > U.S. Cyber Command > News

editions, certifying mission teams), while Cyber Guard illustrates an integration-focused approach that stresses broader operational alignment and readiness.

1.2.1.8 *U.S. CISA Cyber Storm*

"**Cyber Storm**"[34] (often mistakenly written as "CIA Cyber Storm") is a long-running **U.S. national-level cybersecurity exercise series** sponsored by the **Cybersecurity and Infrastructure Security Agency (CISA)**[35] within the Department of Homeland Security. It is designed to bring together **federal agencies, state and local partners, critical infrastructure operators, and selected international participants** to simulate and coordinate the discovery of, and response to, a **nationally significant cyber incident**. Cyber Storm is explicitly focused less on hands-on technical exploitation and more on **policy, procedures, information sharing, coordination mechanisms, and decision-making** under time pressure ,i.e., it exercises the "system of systems" that must function during a large cyber crisis.

From an exercise-typology perspective, Cyber Storm is typically described as a **distributed, operations-based exercise**. Recent iterations have been implemented as **functional exercises** (and in some contexts described as full-scale/distributed formats) where many organizations play from their own locations, using their real incident response structures and communication channels, while an exercise control function injects scenario developments and collects observations. CISA materials emphasise that the series is used to **stress-test** response capabilities "absent the consequences of a real-world event," often using **multi-sector critical infrastructure** scenarios and concurrent incident threads to expose interdependencies and coordination challenges.

Cyber Storm is also closely linked to formal evaluation and improvement practice. CISA positions the series as a national mechanism to examine how well participants' plans, policies, and coordination procedures work at scale, and it produces **After-Action reporting** that consolidates significant findings and improvement themes for both government and participating stakeholders. These reports typically document purpose, scope, planning approach, scenario design, and major observations, thereby supporting institutional learning across exercise cycles rather than treating each exercise as a standalone event.

1.2.1.9 *KYHA – Finnish national cyber exercises*[36]

JYVSECTEC (Jyväskylä Security Technology) is a cybersecurity research, development, training, and exercising centre hosted by **JAMK University of Applied Sciences**. It functions as a **national capability hub** that combines applied research and development with operational training and exercise delivery, and it is closely associated with Finland's national cyber range environment, the **Realistic Global Cyber Environment (RGCE)**. RGCE is described as a cyber range that brings together realistic global and organizational environments in an isolated setting in order to enable safe, state-of-the-art cyber training and exercises under simulated attack conditions. In the Finnish context, JYVSECTEC is particularly significant because it acts not only as a technical platform provider, but also as an organiser and long-term enabler of **recurring national cyber exercises**. This makes it a useful reference point for dissertation work on national cyber exercise ecosystems, as it illustrates how an exercise capability can be institutionalised through a stable, specialised provider while remaining connected to public-sector preparedness goals and whole-of-society resilience.

[34] Cyber Storm: Cyber Storm: Securing Cyber Space | CISA
[35] CISA: Home Page | CISA
[36] Part of the information in this subsection is based on a presentation on JYVSECTEC delivered by Karo Saharinen in February 2026. The presentation was held in the context of the 2026 National Cyber Exercise for Municipalities and Critical Infrastructure Operators (KYHA), conducted on 9–13 February 2026 at the Institute of Information Technology of JAMK University of Applied Sciences, with particular emphasis on strengthening cyber expertise in the logistics sector.

A particularly important element of this Finnish model is the recurring national exercise portfolio commonly referred to as the **KYHA exercises**. **JYVSECTEC states that it conducts four national cybersecurity exercises per year.** These exercises target a broad set of stakeholders, including **municipalities, critical infrastructure organizations, government administration bodies, security authorities, and healthcare organizations**. The exercises are enabled by **Finland's Ministry of Transport and Communications**, while the **Security Committee** participates in a coordinating or guiding role. This governance arrangement is analytically important because it demonstrates how a **national cyber exercise programme** can be anchored in public oversight without requiring the state to build and operate the entire technical capability itself. Instead, Finland appears to rely on a specialised exercise centre with technical, educational, and operational expertise, while strategic alignment and national coordination remain tied to state-level institutions. For research on cyber exercise governance, this is a highly relevant model of **semi-centralised institutionalisation**: a recurring national exercise capability supported by a stable provider, embedded in a broader national resilience architecture, and aligned with public preparedness objectives.

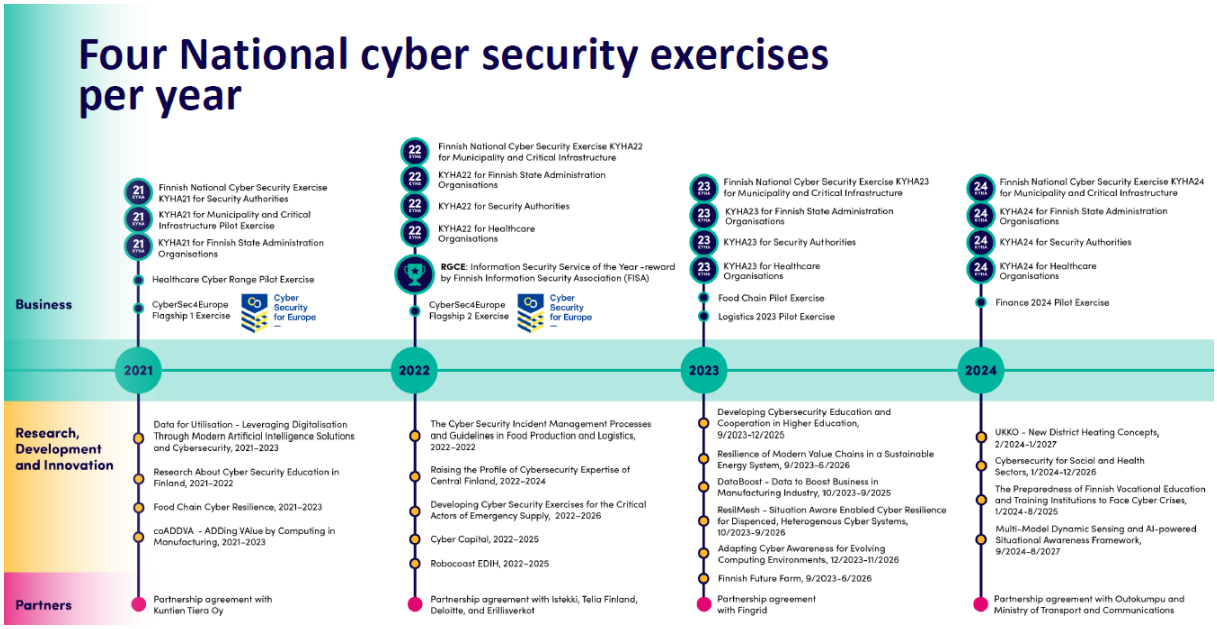


FIGURE 11 - KYHA NATIONAL CS EXERCISES (SOURCE: JYVSECTEC PRESENTATION)

Part of the information in this subsection is based on a presentation on JYVSECTEC delivered in **February 2026** in the context of the **2026 National Cyber Exercise for Municipalities and Critical Infrastructure Operators**, held on **9–13 February 2026** at the **Institute of Information Technology of JAMK University of Applied Sciences**. The **2026 iteration placed particular emphasis on strengthening cyber expertise related to the logistics sector**.

The presentation adds important practical depth to this structural description by showing how the Finnish model works in operational terms. JYVSECTEC originated in **2011** as an EU-funded research and development initiative and gradually developed into a platform that combines education, applied research, and business-to-business cyber exercise services. This historical evolution is highly relevant, because it shows that the Finnish capability did not emerge fully formed as a national programme, but was built incrementally through research funding, infrastructure development, institutional partnerships, and repeated exercise practice. Over time, JYVSECTEC evolved from an academic innovation project into a mature exercise actor with both national and international relevance. From a dissertation perspective, this supports the argument that

sustainable national cyber exercise ecosystems are often the result of **long-term capability-building processes** rather than one-off policy decisions.

The **KYHA exercises** can be characterised as **technical-operational national cyber exercises** intended to prepare **public authorities and essential-service providers** for **complex and large-scale cyber incidents**. JAMK explicitly frames KYHA as an exercise format that supports preparedness for difficult cybersecurity incidents, thereby aiming to strengthen **shared situational awareness, cross-organizational coordination, and practical response capability**. This is significant because it suggests that the Finnish national exercise logic is not limited to isolated organizational drills or abstract awareness-building. Instead, it is designed to test how multiple actors respond under technically realistic conditions when **essential digital services and cross-sector dependencies** are under pressure. Public communications around KYHA also indicate that participation may include both **technical personnel and managerial or communication functions**. This reinforces the view that national cyber exercises must address not only technical detection and response, but also **leadership decision-making, crisis communication, stakeholder coordination, and multi-level incident management**.

A central feature of the Finnish approach is that JYVSECTEC's work should not be seen as mere simulation in the simplified sense. According to the presentation, only the narrative frame of the exercise is fictional, whereas the **technical effects are real within the contained cyber range environment**. Real software is attacked, genuine attack tools are used, and systems fail or behave abnormally in technically credible ways. The cyber range therefore provides a controlled but realistic environment in which participants must respond to consequences that resemble those of real cyber incidents. This distinction is crucial for dissertation analysis because it highlights a **hybrid exercise model**: one that remains safe and isolated, yet still preserves the realism needed for meaningful training, operational stress, and inter-organizational coordination. It also demonstrates the value of a **dedicated national exercise infrastructure such as RGCE**, which enables technically realistic repetition without endangering production environments.

Another particularly valuable characteristic of the Finnish model is its focus on **societal and sectoral interdependencies**. Rather than concentrating only on generic attack scenarios, JYVSECTEC appears to design exercises around the actual digital backbone of Finnish society. The presentation provides several examples illustrating this. In the healthcare context, the range includes **synthetic citizens** generated by the organisers, complete with artificial social security identifiers, because many public-sector processes in Finland depend on such identity-linked records. This makes it possible to create legally safe but operationally realistic healthcare scenarios. In one example, an attack disrupted the data flow between healthcare systems and pharmacies, resulting in unreadable or incorrectly displayed prescriptions. In another, digital monitoring values in a healthcare environment became unreliable, forcing personnel to question whether they could continue trusting system-generated data or whether they had to switch to manual procedures. These examples are especially important for thesis work because they demonstrate that advanced cyber exercises can target not only systems, but also **trust in digital information, continuity of public services, and socio-technical dependencies between institutions**.

The same design logic extends beyond healthcare. In the municipality context, the presentation describes exercise scenarios involving Finland's **highly digitalised property ownership system**, where uncertainty over digital records could create confusion about who owns apartments, houses, or land plots. In the education context, it refers to incidents involving student management systems and the disclosure of transcript data. These examples suggest that the Finnish approach conceptualises cyber resilience as the ability to preserve the **trustworthy functioning of digitally mediated societal services**. For dissertation purposes, this is highly relevant because it broadens the understanding of cyber exercises from classic incident-response drills toward **resilience testing of critical administrative, civic, and economic processes**.

From the perspective of exercise engineering, the presentation also describes a structured methodology for **scenario and attack design**. Each exercise has its own **sector-specific scenario**, its own **attack chains**, and its own **strategic threat logic**. Importantly, the attack chains are documented in advance and may begin at different points in the **MITRE ATT&CK sequence**, rather than always starting with initial reconnaissance. This means the organisers deliberately choose the most pedagogically and operationally relevant entry point for each exercise. In addition, attacker motives such as extortion or other strategic objectives are built into the scenario design. This is methodologically significant because it shows that the Finnish model does not simply attempt to replicate full attack chains for their own sake; rather, it uses carefully selected stages and threat motivations to create specific **pressure points, learning outcomes, and decision-making challenges** for participants.

A further important strength of the Finnish model is that exercises are **not confined to single organizations**. The presentation indicates that organizations often participate together with **partner entities, outsourced service providers, or other interconnected actors**, reflecting the practical reality that no modern organization operates in isolation. Critical systems may be owned, maintained, or supported by external providers, and meaningful incident response therefore requires coordination across organizational boundaries. This is highly relevant for a dissertation on national cyber exercise systems, because it supports the view that national preparedness cannot be meaningfully tested through isolated entity-level drills alone. Instead, cyber exercises must increasingly be understood as **ecosystem exercises** that examine dependencies, communication channels, escalation interfaces, and collaborative response across institutional networks. The KYHA model is therefore particularly instructive as an example of **recurring exercises targeted at different communities while retaining a common national preparedness logic**.

The technical maturity of the exercise environment is another notable aspect. The blue-team side of the cyber range includes a wide range of enterprise-like components such as **web servers, mail services, helpdesk functions, extranets, domain controllers, intranet services, certificate authorities, logging and SIEM-related capabilities, monitoring functions, and file capture for traffic analysis**. Virtual environments can be deployed on demand, and the infrastructure supports **large numbers of concurrent virtual machines**. The broader JYVSECTEC presentation further shows that the range is capable of representing a wide set of organizational environments, including **commerce and financial organizations, critical infrastructure organizations, internet service providers, cloud service providers, logistics supply chains, electricity companies, software companies, healthcare cybersecurity value chains, research facilities, security operations centres, food production value chain organizations, and government organizations**. This breadth suggests that realism in the Finnish model is not only narrative or procedural, but also deeply technical and sector-specific. At the same time, the environment remains isolated from the public internet for legal and safety reasons, especially when potentially dangerous software or attack code is used. The Finnish case therefore illustrates a core design principle of cyber ranges: **realism must be balanced with containment**, and safe exercise execution requires both technical sophistication and strong environmental control.

The presentation also raises an issue that is highly relevant for future-oriented research: the difficulty of reproducing **realistic background behaviour** in cyber ranges. While attack logic can be scripted and systems can be modelled, realistic human-like user behaviour and natural network traffic are much harder to emulate. Earlier work involved bots intended to mimic end-user actions, but the unpredictability of actual human behaviour makes this challenging. This insight is especially important for dissertations dealing with realism, scenario validity, or AI-supported cyber range enhancement, because it shows that exercise fidelity depends not only on attack realism, but also on **plausible ambient behaviour, timing, operational tempo, and socio-technical interaction patterns**. In other words, a realistic cyber exercise environment must simulate not only hostile actions, but also the **normal digital life of an organization or society**.

The broader institutional material further confirms the **maturity and scale** of the Finnish model. JYVSECTEC presents itself as **Finland's leading research, development, and training centre focusing on cybersecurity, artificial intelligence, and data analytics**. Its activities are structured across **three main pillars**: **research and development**, **business services including training and exercises and the JYVSECTEC Hub concept of cyber range**, and **higher education**. The material also provides useful scale indicators for the national exercise ecosystem. It reports **28 KYHA exercises in total**, including **19 conducted between 2021 and 2025**, with **75 participating organizations overall**, **55 of them between 2021 and 2025**, and approximately **3,100 participants in total**. In addition, the presentations report **substantial long-term funding** both for KYHA delivery and for the development of JYVSECTEC exercise activities since 2011. These figures are highly relevant for dissertation analysis because they show that the Finnish model is not an ad hoc initiative, but a **long-term, significantly developed national capability** with continuity, repeated stakeholder participation, and measurable institutional scale.

The threat model described by JYVSECTEC also deserves attention. The organization presents its exercise methodology as a **threat-driven approach** in which cyberattacks are designed to simulate **attack vectors and threat actors together with their tactics, techniques, and procedures**. Exercises begin with **background stories for the training audience** so that the cyber scenarios align with the storyline and risk landscape of the participating organization. The presentations further list a broad range of simulated actors, including **human errors, configuration errors, script kiddies, insiders and disgruntled employees, criminals, hacktivists, activists, and nation- or state-sponsored actors**. The simulated attack spectrum includes **distributed denial-of-service attacks, botnets, ransomware, phishing campaigns, watering hole attacks, malware, website defacements, steganography, advanced persistent threat campaigns, and Trojan horses**. This is especially relevant for dissertation purposes because it provides a clear conceptual basis for describing the Finnish exercise model as **threat-driven, scenario-based, and closely aligned with sectoral risk landscapes**.

The governance model behind the Finnish exercises also offers important lessons. While some national-level metrics and KPI-type approaches exist, JYVSECTEC as a university-based actor cannot impose a mandatory performance regime on all participating organizations, especially in the private sector. Instead, it can recommend metrics, provide reporting, and indicate comparative performance without formally grading participants. This is a useful governance insight: national cyber exercise systems may benefit from **recurring formats, shared methods, and common infrastructure** without necessarily requiring a rigid centralised scoring model. For dissertation purposes, this suggests that national exercise coherence can be achieved through **stable institutional roles, repeatable exercise cycles, and common learning frameworks**, even when performance management remains partly decentralised.

Another major strength of JYVSECTEC is its integration of **education, workforce development, and operational exercise delivery**. Bachelor's and master's degree students may participate in the same cyber exercise courses but with different roles and objectives. Bachelor's students often focus more on technical defensive tasks, while master's students may be given broader coordination and management roles. Some students may play threat actors, while the **white team** orchestrates the scenario, controls timing, supervises safety, and authorises cyber effects. This layered model demonstrates that the exercise ecosystem serves multiple purposes simultaneously: **professional education, capability development, operational training, and national preparedness support**. From a research perspective, this is especially relevant because it shows how a national exercise capability can contribute to resilience not only through single events, but also through **sustained talent development and the creation of an exercise-experienced workforce**.

Finally, the Finnish case also appears important as an **innovation platform**. The presentations refer to continuing investment in **local AI infrastructure** and discuss the possibility of incorporating intentionally unreliable or biased AI into exercise environments, for example scenarios in which

participants must decide whether to trust or distrust AI-generated support. They also mention participation in **large European defence-related cyber range initiatives**. This indicates that JYVSECTEC is not only a provider of current exercise services, but also a site of experimentation in which future cyber and socio-technical threats can be explored. For dissertation work on next-generation cyber exercises, this is highly significant because it suggests that mature cyber exercise ecosystems may increasingly function as **laboratories for testing not only cyber incidents, but also the organizational, ethical, and operational implications of emerging technologies such as AI**.

From the perspective of cyber exercise design and governance, at least **three transferable lessons** emerge from the Finnish model. First, JYVSECTEC demonstrates the value of a **dedicated national exercise infrastructure such as RGCE**, which enables **repeatability, technical realism, and safe simulation** without endangering production systems. Second, the **KYHA portfolio** illustrates how national cyber exercises can be **institutionalised as a recurring programme with multiple iterations per year** and different target communities, thereby supporting **cumulative maturity building and cross-cycle learning**. Third, the governance and participation model highlights the importance of **public-private and cross-authority coordination**: ministries, national coordinating bodies, municipalities, critical infrastructures, and other stakeholders are all part of the preparedness logic, making **interdependency management and collaborative response** central exercise objectives.

Overall, the Finnish model centred on **JYVSECTEC**, **RGCE**, and the recurring **KYHA exercises** represents a comparatively advanced example of a **national cyber exercise ecosystem**. It combines **stable institutional hosting, technical-operational realism, recurring national exercise delivery, sector-specific scenario engineering, multi-actor participation, and strong linkage between research, training, and public preparedness**. For a dissertation on cyber exercise governance and national cyber resilience, the case is particularly valuable because it shows how a national programme can be built around a **dedicated exercise centre** that is simultaneously a research actor, a training provider, and an operational exercise organiser. It also demonstrates that the most mature cyber exercise systems move beyond isolated organizational drills and instead test the resilience of **digitally interconnected societal functions, cross-sector coordination mechanisms, and trust in critical information systems under realistic but safely controlled attack conditions**.

1.2.1.10 *LÜKEX – Germany´s national crisis-management exercises*

LÜKEX[37] (short for *Länder- und Ressortübergreifende Krisenmanagementübung*) is Germany's recurring, national-level **strategic crisis-management exercise** format. It is coordinated by the **Federal Office of Civil Protection and Disaster Assistance (BBK)** and conducted jointly with the **federal states (Länder)** and relevant federal ministries, with the explicit aim of improving **whole-of-government crisis management** at the political–administrative level. LÜKEX is designed to test and strengthen how the Federation and the Länder coordinate during complex crises, including the involvement of **aid organizations and critical infrastructure (KRITIS) operators** where relevant to the scenario.

From an exercise-typology perspective, LÜKEX is not primarily a technical "live-fire" cyber range event. Instead, it is a **governance and coordination exercise** that targets decision-making, leadership, communication, and inter-organizational coordination under pressure. The exercise is built around a **fictional national crisis scenario** that provides the common frame for participants to activate their crisis structures, exchange information, coordinate actions, and manage consequences across administrative boundaries. The focus on strategic-level coordination makes

[37] LÜKEX: LÜKEX - BBK

LÜKEX comparable to other national crisis-management exercise programmes where the main training audience is the crisis governance layer rather than frontline responders.

Although LÜKEX is often perceived through its execution period, the series follows a full exercise lifecycle consistent with good practice in national programmes: (1) **topic selection and scenario development** aligned with relevant national risks and interdependencies; (2) **extensive preparation and coordination** across participating authorities and stakeholders to ensure realistic play, role clarity, and coherent exercise control; (3) **distributed execution** in which crisis staffs and coordination bodies work through the simulated situation in real time; and (4) **structured evaluation** that consolidates observations into lessons and actionable recommendations. This evaluation emphasis is visible in BBK's publication of **Auswertungsberichte (evaluation reports)** for specific iterations, which document objectives, exercise set-up, and key findings for improving German crisis management across cycles.

A recent example illustrates LÜKEX's relevance for cyber-related dissertation work: **LÜKEX 23** (conducted in 2023) used the scenario of a **cyberattack on governmental and administrative functions** ("Cyberangriff auf das Staats- und Regierungshandeln"), explicitly exercising how Bund and Länder maintain the state's ability to act under a nationwide cyber crisis. BBK reports that **over 2,500 participants** from federal and state authorities and other crisis-management organizations took part, underlining the scale and the multi-actor coordination focus of the format.

For dissertation purposes, LÜKEX is a strong reference case for how a national programme can institutionalise **strategic crisis exercising** with an explicit learning loop. Its core contribution is not technical validation of defensive tools, but the systematic testing of **interfaces between authorities**, the creation of shared situational awareness, coherent public-sector decision-making, and the integration of critical infrastructure stakeholders into national crisis governance. In this sense, LÜKEX demonstrates how cyber crises can be exercised as a **state capability challenge**, where the main outcome variables are coordination quality, timeliness of decisions, clarity of responsibilities, and the effectiveness of cross-level communication ,supported by formal evaluation reporting that enables iterative improvement across exercise cycles.[38]

1.2.1.11 *KSÖ Planspiele – Austrian National Cyber Exercise*[39]

The **Kompetenzzentrum Sicheres Österreich**[40] **(KSÖ) Cybersecurity Planspiele** are a recurring Austrian exercise format that brings together stakeholders from **public authorities, critical infrastructure and private-sector operators, and research/technology partners** to rehearse cyber crises under controlled conditions. In Austria's broader preparedness ecosystem, these planspiele serve a dual purpose: they are (i) a learning and coordination instrument for participants and (ii) a practical "stress test" for national cooperation mechanisms that must function during high-impact incidents. The KSÖ itself frames its planspiele as realistic trainings that enable state and industry to act more rapidly and confidently in crises, and as a setting in which participants can test protective measures and support each other.

A defining feature of the KSÖ planspiele is their explicit focus on **cross-organization coordination**. This is strategically relevant because major cyber incidents rarely remain confined to a single organization. Even if the initial compromise is technical, the incident quickly becomes socio-technical: it creates uncertainty, time pressure, competing priorities between security and continuity, external reporting requirements, and dependency-driven cascading effects. In such

[38] I was participating on the organizer site the LÜKEX in 2023.
[39] I was organizing the KSÖ Planspiele in 2018 and 2021.
[40] KSÖ, Planspiele und Simulationsübungen - Kompetenzzentrum Sicheres Österreich

settings, resilience depends not only on the competence of individual responders but on the ability of different entities to align, operationally and organizationally, under stress.

The planspiel format is particularly suitable for testing these "interface capabilities." It allows participants to practice **escalation, role clarity, joint situational understanding, and coordinated decision-making** without the operational risks associated with touching production systems. At the same time, the repeated nature of the KSÖ planspiele suggests an intent to institutionalise exercising as a sustained capability-development mechanism rather than as a one-off awareness event. For example, Austria's CERT community has documented that the KSÖ conducted its Cybersecurity Planspiel repeatedly and explicitly characterises the 2017 iteration as the fourth occurrence, emphasising that it was used to exercise Austria's "fitness" against cyberattacks, espionage, and online fraud.[41]

Evolution and continuity of the exercise series

Public sources indicate that the KSÖ planspiele have evolved over time both in **scale** and in **technical realism**. A widely referenced milestone is the **2017 Cybersecurity Planspiel (6–7 November 2017)**, described by multiple sources as the fourth iteration of this national exercise series. Austria's CERT reporting highlights the recurrence and the ambition of stress-testing national preparedness. The Austrian Federal Computing Centre (BRZ) likewise notes participation and frames the planspiel as a "stress test" of national fitness against cyberattacks and related threats[42]. Complementing these institutional statements, the AIT (Austrian Institute of Technology[43]) describes the 2017 event as a large-scale exercise hosted on its premises, stating that around **200 experts** tested crisis response to cyberattacks on critical infrastructures, illustrating both the scale and the intent to simulate realistic conditions.

The series later expanded into **cross-border participation and hybrid delivery**. For the **KSÖ Planspiel 2021**, the AIT project description states that the exercise was realised on the **AIT Cyber Range CaSTLE** and explicitly included remote integration of teams from **Germany and Switzerland**, reflecting a deliberate move toward multi-country participation and technically supported execution.[44] Austria's official online safety portal similarly describes a "DACH" cross-border exercise format and notes a focus on cyber-physical challenges and accompanying information measures, which points to a scenario framing that integrates both operational effects and the information dimension of crisis management.

More recently, public reporting confirms continuity and further scaling. Austria's Ministry of the Interior (BMI) reports that the KSÖ conducted its **seventh Cyber Planspiel in November 2024** in cooperation with AIT and an industry partner, describing it as a fictitious hybrid attack on state, economy, and society executed on modern IT infrastructure and involving Austrian and international organizations. In the same timeframe, KSÖ communications describe a scenario in which participants act in different roles within fictional organizations affected by a large cyberattack wave, again underlining role-play, coordination, and decision-making under stress as central mechanisms.[45] Public announcements also show a continued emphasis on multi-stakeholder and international participation, naming involvement from multiple countries and including government entities and CERT communities.

Taken together, these elements suggest a maturation trajectory that is typical of national exercise ecosystems: early iterations build shared language and trust; later iterations increase realism

[41] **KSÖ Cybersecurity Planspiel 2017,** https://www.cert.at/de/berichte/2017/2017-chap03-4
[42] Sicherheit im Netz - BRZ nimmt am KSÖ-Cybersecurity-Planspiel 2017 teil, Sicherheit im Netz - BRZ nimmt am KSÖ-Cybersecurity-Planspiel 2017 teil - BRZ
[43] AIT , Startseite - AIT Austrian Institute Of Technology
0 [44] **Planspiel-KSÖ 2021**, https://www.ait.ac.at/themen/cyber-security/projects/planspiel-ksoe-2021
[45] Planspiele und Simulationsübungen, https://kompetenzzentrum-sicheres-oesterreich.at/projekte/planspiel/

through cyber range support, larger multi-entity participation, and scenario designs that incorporate hybrid and cross-domain aspects.

Exercise format and methodological characteristics

Although public sources do not always disclose full internal exercise artefacts, the available descriptions allow several methodological characteristics to be inferred with reasonable confidence.

(1) Scenario-driven, role-based crisis simulation: KSÖ describes its planspiele as realistic crisis training in which participants "test suitable protective measures" and "support each other," implying that the setting is designed for decision-making, coordination, and joint problem solving rather than for isolated technical tasks. The 2024 description explicitly references role allocation within fictional organizations and a large cyberattack wave, which is consistent with structured role-play to make decision and coordination behaviour observable.

(2) Increasing technical grounding through cyber range infrastructure: The AIT's project description for 2021 clearly states that the planspiel was implemented and executed on the AIT Cyber Range CaSTLE. This indicates that at least some iterations go beyond purely discussion-based tabletop methods and incorporate technically grounded simulation environments, which can improve realism and, crucially, enable stronger evidence capture (e.g., logs, event traces, timed inject delivery). The AIT positions its cyber range as a safe and realistic environment to simulate critical systems and test incidents without affecting production systems, aligning with the requirements of critical infrastructure exercising.

(3) Scale and multi-entity coordination as a design goal: The 2017 sources consistently emphasise that the exercise is multi-entity and large-scale (e.g., the "fourth time," national fitness testing; participation by major actors; AIT describing around 200 experts). Such scale changes the nature of what is exercised: it shifts focus from individual technical competence to ecosystem coordination, information sharing, escalation, cross-organizational dependencies, and aligned crisis communication.

(4) Hybrid-threat framing and information dimension: In 2024, the BMI describes a fictitious "hybrid attack" on state, economy, and society. The 2021 DACH exercise description also highlights cyber-physical challenges and accompanying information measures, pointing to the integration of technical disruption with communication and influence dimensions. This is consistent with contemporary national-security-oriented exercising, where cyber incidents are rarely purely technical and often interact with disinformation, public perception, and strategic messaging.

Another relevant characteristic is the role of **technology and research partners** in enabling the exercise environment. In public materials, AIT is repeatedly referenced as an implementation partner and host, including providing the cyber range platform used for exercise execution. The 2024 BMI reporting explicitly mentions cooperation between KSÖ and AIT, alongside an industry partner, which reflects an ecosystem approach: governance and convening through KSÖ, technical exercise infrastructure through a national research actor, and sector realism through private operators.

This pattern matters for national exercise governance. Large-scale exercises generally require a division of labour: scenario and inject design; technical environment provisioning; control and evaluation functions; and stakeholder coordination. Where this is institutionalised through partnerships, exercising becomes more sustainable and can scale across years.

Relevance for national cyber resilience governance

From a national governance perspective, the KSÖ planspiele illustrate several design lessons that are broadly transferable beyond Austria. First, they demonstrate the value of a **recurring national exercise instrument** that builds continuity and community memory. Repetition enables

progression: exercises can revisit known interface issues (e.g., escalation, information sharing, crisis communication) and verify whether improvements have actually been implemented.

Second, the evolution toward **cyber range–supported execution and hybrid formats** provides a pathway to balance inclusiveness with realism. Tabletop- and role-based elements enable broad participation across functions and organizations, while range-supported components allow more technical grounding and stronger evidence capture where necessary.

Third, the emphasis on **multi-entity participation** directly addresses one of the hardest aspects of cyber resilience: interdependency management and coordination under uncertainty. Exercises that bring together authorities, operators, and supporting organizations create a controlled environment to test whether collaboration mechanisms function under stress, not just whether individual organizations have internal procedures.

Finally, the public descriptions suggest that KSÖ planspiele increasingly aim to approximate **national-scale stress conditions**, large attack waves, hybrid dimensions, cross-border participation, rather than limiting scope to isolated incident narratives. This is aligned with the governance reality that national resilience depends on the ability to manage concurrent incidents, cascading effects, and multi-stakeholder information demands.

Limitations of public visibility and implications for research

At the same time, publicly available information about the Planspiele tends to focus on participation, general goals, and high-level descriptions, while detailed information about **exercise artefacts, evaluation criteria, metrics, and longitudinal improvement tracking** is less visible. This is understandable given sensitivity and confidentiality considerations, but it has implications for national programmes that aim to demonstrate measurable resilience improvement: national value increases when exercises generate structured, comparable artefacts (e.g., objective mapping, inject logs, evidence plans, after-action outputs) and when improvements are tracked and re-tested in later cycles.

For this dissertation's broader argument, focused on repeatability, comparability, and auditability, KSÖ Planspiele provide a concrete national example of how a country can institutionalise recurring cyber crisis exercises and progressively increase fidelity and scope through partnerships and cyber range support. The case also motivates why structured artefact standards and auditable workflows matter: they are the mechanism by which recurring exercises can be turned into defensible national learning and measurable improvement, even under confidentiality constraints.

Summary of International Cyber Exercise Landscape

The analysis of the international cyber exercise landscape demonstrates that cyber exercising has evolved into a diverse ecosystem of organizational, sectoral, national, European, international, civilian, and military practices. Cyber exercises range from executive discussion-based formats to technically intensive live-fire simulations, distributed multinational exercises, national crisis-management exercises, and threat-led resilience testing. This diversity reflects the fact that cyber resilience itself is multidimensional. Technical defence, crisis coordination, executive decision-making, continuity, communication, legal response, cross-organizational cooperation, and military mission performance cannot all be validated through a single exercise format.

The international landscape therefore confirms the fragmentation component of H1. Considerable differences remain in terminology, exercise typologies, institutional ownership, planning methodologies, technical environments, documentation practices, evaluation approaches, and the extent to which exercises are connected to formal improvement mechanisms. The term *cyber exercise* consequently refers to activities with substantially different purposes and evidentiary value. The mere fact that two organizations or countries conduct cyber exercises does not mean that their exercise programmes, results, or levels of preparedness can be meaningfully compared.

At the same time, the comparative analysis demonstrates that fragmentation does not mean complete methodological incompatibility. Across otherwise different exercise families, recurring analytical characteristics can be identified. These include clearly defined institutional ownership, explicit exercise objectives, a defined scope, identifiable exercise formats and capability levels, scenario-based execution, structured roles, evaluation activities, and some form of post-exercise learning. Recurring national and international programmes additionally demonstrate the importance of institutional continuity, repeated participation, cross-organizational coordination, and the possibility of comparing findings across exercise cycles.

A particularly important conclusion is that the field does not require one uniform cyber exercise model in order to improve comparability. Different exercise objectives necessarily require different formats. Technical cyber defence, executive crisis-management, national coordination, and multinational interoperability cannot be reduced to one standard exercise structure without losing operational relevance. The more appropriate goal is therefore to establish common principles for describing, governing, documenting, and evaluating exercises while preserving flexibility in scenario design and execution.

This observation is central to the dissertation's distinction between **convergence and harmonisation**. The emerging pattern should not be interpreted as evidence that national, sectoral, or organizational exercise practices are becoming identical. Rather, the international landscape suggests the possibility of convergence around a shared baseline of quality and governance characteristics while diversity in implementation remains. The following chapters investigate this proposition further by examining whether established international methodologies, standards, and European regulatory developments reinforce the same recurring principles.

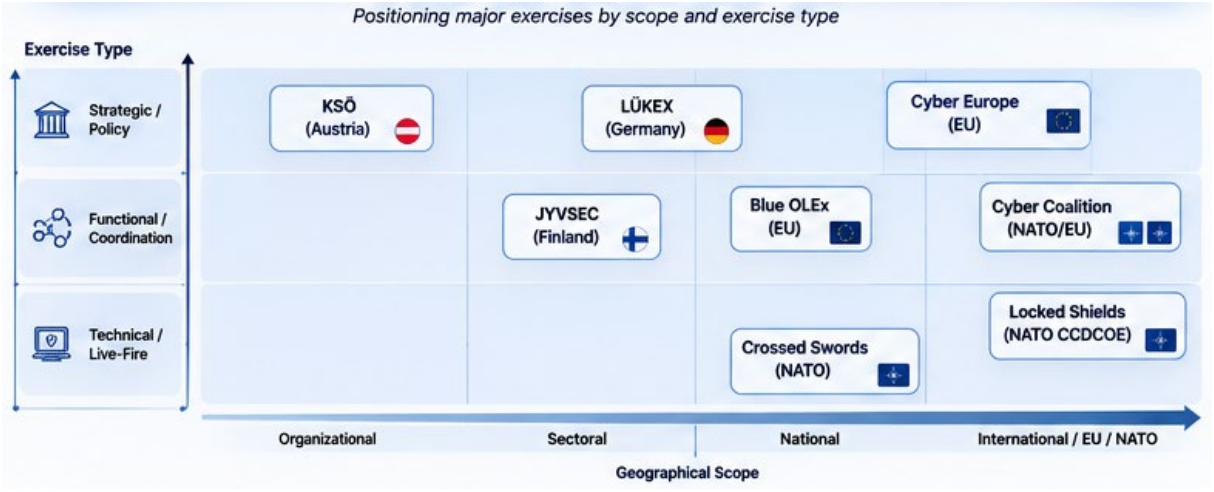


FIGURE 12 - INTERNATIONAL CX LANDSCAPE - SELECTED REFERENCE EXERCISES

The findings of Section 1.2 therefore provide the empirical foundation for Chapter 1.3. The international landscape identifies the diversity of existing practices and the analytical dimensions necessary to describe them consistently. Chapter 1.3 subsequently examines established exercise methodologies and standards, including HSEEP, ENISA, and BSI, to determine whether these frameworks contain recurring lifecycle, governance, documentation, evaluation, and improvement principles. This allows the dissertation to move from observing similarities across individual exercises to analysing whether methodological convergence can be demonstrated more systematically.

The same analytical logic is subsequently extended in Chapter 1.4 to the European regulatory and policy environment. The purpose is to determine whether European resilience requirements further reinforce expectations concerning governance, preparedness, continuity, dependency management, testing, documentation, evidence, and improvement. In combination, the

international exercise comparison, methodology analysis, and regulatory analysis provide the evidence base from which the **European Convergence Framework for Cyber Exercises** is derived in Chapter 1.5.

The international exercise landscape therefore supports an important preliminary conclusion for H1. Contemporary cyber exercise practice remains heterogeneous, and substantial differences continue to exist between countries, sectors, institutions, and exercise families. Nevertheless, a set of recurring structural characteristics can already be observed. These characteristics suggest that convergence is more likely to occur at the level of **governance principles, lifecycle discipline, artefact structure, evaluation, evidence, dependency awareness, and continuous improvement** than at the level of identical exercise formats.
For the dissertation's broader research logic, this conclusion has two consequences. First, it provides the empirical basis for the European Convergence Framework developed later in Part I. Second, it establishes the analytical foundation for the formal Cyber Exercise Taxonomy and Minimum Dataset developed in Section 2.3. The taxonomy does not replace the diversity identified in this chapter. Rather, it translates that diversity into a structured information model through which heterogeneous exercises can subsequently be governed and compared at national programme level.
The international landscape can therefore be summarised through a central finding: **cyber exercise diversity is both necessary and persistent, but diversity does not preclude comparability when exercises are described and evaluated through common analytical dimensions**. This finding provides the transition from the descriptive international landscape to the methodological and regulatory convergence analysis that follows.

1.3 International standards, norms, and best practices for cyber exercises

In today's digital landscape, organizations face an expanding number of cyber risks. Cybersecurity exercises help prepare, test, and enhance team and system skills to respond to attacks. A well-structured cybersecurity exercise methodology provides a holistic approach to simulation and training, increasing resilience and agility in addressing cyber risks.

This chapter and their subchapters give an overview of the identified lessons learned, best practices at the European level followed by the relevant regulations and standards.



FIGURE 13 - CONVERGENCE OF EXERCISE METHODOLOGIES

The figure shows that major international approaches such as HSEEP, ENISA, and BSI follow a broadly similar exercise lifecycle, moving from initiation and design through preparation, execution,

evaluation, and improvement. Despite methodological differences, they share core quality principles such as objective-driven planning, clear governance, structured artefacts, evidence-based evaluation, after-action reporting, continuous improvement, proportionality, and human oversight.

Homeland Security Exercise and Evaluation Program (HSEEP)[46]



FIGURE 14 - HSEEP EXERCISE CYCLE

The Homeland Security Exercise and Evaluation Program (HSEEP) is the U.S. government's "common methodology" for planning, running, and judging homeland security and emergency management drills. The main goal is to make exercises across businesses more similar, repeatable, and focused on improvement by providing common language, standardized processes, and uniform documentation throughout the entire exercise cycle, from program management to evaluation and corrective actions.

HSEEP sees exercises as part of a system for continuous improvement. Exercises test plans and abilities in controlled settings, observations are recorded using structured evaluation methods, and the results are turned into an After-Action Report (AAR) and an Improvement Plan (IP) that lists who is responsible for what and when. HSEEP makes it clear that post-exercise activities like analysis, prioritization, and tracking are necessary to get value, not only the exercise itself.

HSEEP offers a building-block approach that goes from cheaper, discussion-based activities to more accurate operational validation. It makes a clear distinction between discussion-based exercises (like seminars, workshops, tabletop exercises (TTX), and games) and operations-based exercises (which focus on real-time coordination and response performance under simulated conditions and need more control and logistics, like controllers, simulation cells, and actors/simulators).

HSEEP encourages clearly defined exercise goals and roles (like planners, facilitators/controllers, evaluators, and players), standardized data collection during play (including structured debriefs/hot washes), and formal documentation products (like scenario and event documentation, AAR/IP). It also knows that some exercise artifacts, such scenario data, AAR/IP, and event lists, may be sensitive and should be treated that way.

Furthermore, the HSEEP website contains a bright palette of templates from 2020. These templates could be also used as a basis for the AI-enhanced CX Lifycycle.

Beyond its role as an exercise management methodology, HSEEP is also relevant for this dissertation because its lifecycle structure provides a practical basis for organising evaluation data and metrics across programme management, exercise design and development, exercise conduct, and exercise evaluation and improvement planning.

[46] HSEEP, About - HSEEP Resources - Preparedness Toolkit, Accessed on 21-02-2026. Note: Since the government has stopped funding the website, it will not be regularly managed. The last time this website was changed was on February 13, 2026. It won't be changed again until funding is in place.

ENISA

ENISA, the European Union Agency for Cybersecurity, works to make sure that all of Europe has a high level of cybersecurity. The European Union Agency for Cybersecurity was set up in 2004 and strengthened by the EU Cybersecurity Act. It helps shape EU cyber policy, makes ICT products, services, and processes more trustworthy through cybersecurity certification schemes, works with Member States and EU bodies, and gets Europe ready for the cyber challenges of the future. The Agency works with its main partners to build trust in the connected economy, make the Union's infrastructure more resilient, and, in the end, keep Europe's society and citizens safe online. They do this by sharing knowledge, building capacity, and raising awareness.

ENISA's exercise lifecycle best practices can be summarised as a set of repeatable design principles applied across six phases (Initiation → Design → Preparation → Execution → Evaluation → Moving Forward). The methodology is explicitly meant to be end-to-end, supported by templates and checklists, and it uses go/no-go checkpoints between phases to reduce planning oversights and prevent "running an exercise that is not ready to run.[47]

In the Initiation phase, ENISA's best practice is to start with a clearly articulated purpose ("the why") and then select an exercise type that realistically matches that purpose and the available maturity and resources. A structured feasibility assessment is recommended to confirm alignment with strategic goals, risk priorities, and legal/compliance needs, to secure stakeholder commitment, and to ensure technical and organizational readiness. ENISA also recommends assessing the current cybersecurity posture (e.g., via gap analysis) so that exercise objectives target real weaknesses rather than assumptions.

In Design, ENISA stresses that exercises succeed or fail on objective quality. Objectives should be SMART and mapped to relevant objective themes (strategic, operational, human factors, process improvement, business continuity, and compliance/legal). ENISA further recommends defining scope explicitly across systems/infrastructure, organizational boundaries (including external partners), threat scenarios, and time/operational constraints, and to revisit scope when major changes occur (e.g., new critical vulnerabilities or organizational changes). Early communication planning is also treated as a best practice to build buy-in and keep key stakeholders aligned throughout the lifecycle.

In Preparation, ENISA highlights three best practices: rigorous scenario engineering, evaluation design "from the start," and stakeholder readiness. Scenarios should be operationalized through a Master Scenario Event List (MSEL) that structures events, incidents, and injects, enabling controlled delivery of information and actions during the exercise. Evaluation should be planned before execution through an evaluation plan that defines roles (evaluators, evaluated teams, report recipients), capability targets (especially for larger exercises), data collection tools, and timelines. Stakeholders should be prepared with clear roles, expectations, and practical guidance to reduce confusion during play.

For Execution, ENISA frames best practice as disciplined orchestration and evidence capture: run pre-exercise checks, execute the scenario with active monitoring to keep flow realistic and aligned with objectives, and perform immediate post-exercise activities (e.g., debriefings) while observations are fresh. This phase is treated as both a performance event and a structured data-collection opportunity for later analysis.
In Evaluation, ENISA places emphasis on producing an actionable After-Action Report (AAR) and controlling sensitivity using the Traffic Light Protocol (TLP). The AAR typically contains an executive summary, exercise overview, methodology, findings/lessons identified, and "moving forward" recommendations; internal AAR versions can include more detailed metrics and ratings, while external versions can be summarised to protect sensitive details. ENISA also explicitly warns against common analysis pitfalls (e.g., over-analysing small datasets) and encourages evaluation methods that prioritise meaningful insights over statistical complexity.

[47]ENISA CX Methodology: https://www.enisa.europa.eu/publications/the-enisa-cybersecurity-exercise-methodology

Finally, in Moving Forward, ENISA treats improvement as the endpoint: exercise value depends on dissemination and action. Best practice includes tailoring dissemination products to audiences (players, planners, the wider organization, management, and ,when relevant ,the public) and converting findings into a structured action plan with prioritisation, scope and ownership assignment (e.g., using a RACI logic), concrete tasks and deadlines, and follow-up verification. ENISA distinguishes between (i) findings about organizational capabilities and (ii) findings about exercise design/execution quality, and recommends feeding both into the organization's capacity-building programme; a replay or follow-up exercise is highlighted as a good practice to verify remediation effectiveness.

1.3.1.1 *ENISA Good Practices on National Exercises*[48]

The ENISA "Good Practice Guide on National Exercises – Enhancing the Resilience of Public Communications Networks" (2009) gives Member States a structured, practice-based way to plan, run, and evaluate national-level exercises that make public e-communications networks and critical information infrastructures more resilient. The handbook is clearly meant to help authorities who haven't done many exercises before, and it fits in with the EU's aim at the time to strengthen prevention, preparedness, coordination, and awareness, with the goal of building toward pan-European exercises.

One important point is that exercises are not just "training events," but also a way to manage resilience and governance in the industry. ENISA says such exercises help authorities and operators focus on specific deficiencies, work together more, find interdependencies, encourage better continuity planning, and build a culture of working together that makes the system more resilient over time. The guide also says that national exercises are especially important because most serious incidents have effects on multiple organizations and can spread to others. For example, problems with communications networks can affect other operators, emergency services, field response, and restoration work in other critical infrastructures. This is why coordinated procedures and trusted communication channels are so important. The guide suggests a distinct four-stage exercise lifecycle: identifying, planning, executing, and assessing.

Organizers should start with the measures and procedures to be tested (including the persons who will be participating in those processes) during the identifying stage. After that, they should choose the type of exercise, the scenario, the participants, and the scope. If you don't have a lot of money, knowledge, or stakeholder commitment, the advice says to start small and work your way up to more complicated formats over time.

ENISA says that planning often takes more time and resources than expected, therefore it needs realistic scheduling, early involvement of stakeholders, and a gradual planning process that creates confidence and consensus. In public-private contexts, recruitment is seen as a major problem. People don't always have to participate, and financial incentives are usually small, so organizers should lower obstacles (particularly worries about privacy) and make the benefits clear.

Trust is often cited as a key factor in success. ENISA says that trust is built through clear planning, strict protection of sensitive information, and repeated collaboration, starting with simple tasks like communication checks and moving on to more difficult ones. The handbook underlines the importance of strict exercise management through a central moderator/director and the

48

employment of monitors who sit near to participant teams, watch what they do for subsequent evaluation, and give scenario injects to keep the play on track while keeping the scenario coherent. ENISA also talks about how important media policy is for planning: organizers should plan beforehand how media will be handled during and after the exercise. Media can also be used as an inject channel to make public information pressure feel more real. ENISA says that the review phase needs to be planned ahead of time and handled carefully to avoid blame and protect partnership. The handbook says that evaluation processes should include everyone and come to an agreement on what needs to be done and what needs to be improved. It should use a variety of sources, such as monitor reports, participant questionnaires, and structured debriefings (such "hot wash-ups" and workshops). To deal with sensitivity, ENISA proposes making many versions of the report for different audiences. For example, there may be a confidential feedback report for stakeholders, a sector-level consensus report that doesn't name organizations, and a public-facing summary that leaves out sensitive information.

Lastly, the book sees evaluation as a way to keep improving: by doing tests again to see how much better things are, by following up with stakeholders to help fix problems, and by keeping a "lessons log" to build up the organization's memory over time.
The book goes beyond national scope and links national exercises to European capability building by saying that communications infrastructures and threats are becoming more cross-border and interdependent. It advises that people gradually strengthen their skills, starting with smaller, discussion-based exercises and working up to transnational or pan-European exercises. This will make it possible for larger-scale cooperation to work.

1.3.1.2 *ENISA Cyber Security Exercise Methodology*[49]

ENISA's exercise lifecycle best practices can be summarised as a set of repeatable design principles applied across six phases (Initiation → Design → Preparation → Execution → Evaluation → Moving Forward). The methodology is explicitly meant to be end-to-end, supported by templates and checklists, and it uses go/no-go checkpoints between phases to reduce planning oversights and prevent "running an exercise that is not ready to run."
In the Initiation phase, ENISA's best practice is to start with a clearly articulated purpose ("the why") and then select an exercise type that realistically matches that purpose and the available maturity and resources. A structured feasibility assessment is recommended to confirm alignment with strategic goals, risk priorities, and legal/compliance needs, to secure stakeholder commitment, and to ensure technical and organizational readiness. ENISA also recommends assessing the current cybersecurity posture (e.g., via gap analysis) so that exercise objectives target real weaknesses rather than assumptions.

In Design, ENISA stresses that exercises succeed or fail on objective quality. Objectives should be SMART and mapped to relevant objective themes (strategic, operational, human factors, process improvement, business continuity, and compliance/legal). ENISA further recommends defining scope explicitly across systems/infrastructure, organizational boundaries (including external partners), threat scenarios, and time/operational constraints, and to revisit scope when major changes occur (e.g., new critical vulnerabilities or organizational changes). Early communication planning is also treated as a best practice to build buy-in and keep key stakeholders aligned throughout the lifecycle.

In Preparation, ENISA highlights three best practices: rigorous scenario engineering, evaluation design "from the start," and stakeholder readiness. Scenarios should be operationalized through a Master Scenario Event List (MSEL) that structures events, incidents, and injects, enabling controlled delivery of information and actions during the exercise. Evaluation should be planned before execution through an evaluation plan that defines roles (evaluators, evaluated teams, report

[49] European Union Agency for Cybersecurity (ENISA). (2026, February 16). *The ENISA cybersecurity exercise methodology: End-to-end guide on how to plan, run and evaluate an exercise* (Version 1.0). Last accessed on February 17, 2026, Link: The ENISA Cybersecurity Exercise Methodology.pdf

recipients), capability targets (especially for larger exercises), data collection tools, and timelines. Stakeholders should be prepared with clear roles, expectations, and practical guidance to reduce confusion during play.

For Execution, ENISA frames best practice as disciplined orchestration and evidence capture: run pre-exercise checks, execute the scenario with active monitoring to keep flow realistic and aligned with objectives, and perform immediate post-exercise activities (e.g., debriefings) while observations are fresh. This phase is treated as both a performance event and a structured data-collection opportunity for later analysis.

In Evaluation, ENISA places emphasis on producing an actionable After-Action Report (AAR) and controlling sensitivity using the Traffic Light Protocol (TLP). The AAR typically contains an executive summary, exercise overview, methodology, findings/lessons identified, and “moving forward” recommendations; internal AAR versions can include more detailed metrics and ratings, while external versions can be summarised to protect sensitive details. ENISA also explicitly warns against common analysis pitfalls (e.g., over-analysing small datasets) and encourages evaluation methods that prioritise meaningful insights over statistical complexity.

Finally, in Moving Forward, ENISA treats improvement as the endpoint: exercise value depends on dissemination and action. Best practice includes tailoring dissemination products to audiences (players, planners, the wider organization, management, and ,when relevant ,the public) and converting findings into a structured action plan with prioritisation, scope and ownership assignment (e.g., using a RACI logic), concrete tasks and deadlines, and follow-up verification. ENISA distinguishes between (i) findings about organizational capabilities and (ii) findings about exercise design/execution quality, and recommends feeding both into the organization’s capacity-building programme; a replay or follow-up exercise is highlighted as a good practice to verify remediation effectiveness.

This methodology will be described later in detail in Chapter ENISA-Based AI-Enhanced Cyber Exercise Lifecycle**,**, and used as the basis for the AI-enhanced lifecycle.

BSI Standards

1.3.1.3 *BSI Standard 100-4 as a methodological backbone for cyber exercises*

The German BSI Standard 100-4 provides a highly operational and governance-oriented approach to emergency management and explicitly positions **tests and exercises** as recurring, structured instruments for preparedness. Although the standard addresses emergency management in a broader sense, its logic transfers directly to cyber crises because cyber incidents increasingly manifest as organization-wide emergencies with impacts on critical processes, service availability, regulatory obligations, and reputation. From a cyber resilience perspective, the value of BSI 100-4 lies in its emphasis on **systematic planning, controlled execution, and continuous improvement**, rather than isolated training events.

Within a cyber context, BSI 100-4 can be understood as a framework that helps organizations move from “having plans” to **proving that plans work** under realistic constraints. The standard encourages a programmatic view: exercises are embedded into an organizational cycle of preparedness, validation, and optimisation. This supports repeatability, learning, and management accountability ,particularly relevant when cyber exercises contribute to cross-entity coordination, national preparedness programmes, or compliance-driven assurance requirements.

Types of cyber exercises in the BSI 100-4

BSI 100-4 supports the idea that organizations should maintain a **portfolio of exercise formats**, selected according to objectives, maturity, risk exposure, and organizational scope. In cyber security, this portfolio approach is crucial because different exercise formats validate different layers of capability ,from governance to operations to technical response.

Discussion-based / Tabletop exercises (TTX)

These exercises focus on leadership and coordination rather than hands-on technical activity. They are used to validate decision-making processes, roles, and communication paths under pressure. Typical objectives include testing escalation thresholds, crisis governance, and the ability to produce an agreed “common operating picture.” Tabletop formats are particularly effective for verifying whether management, IT, security, legal, HR, communications, and business units share the same assumptions and responsibilities in a cyber crisis.

Functional exercises (process and coordination exercises)

Functional exercises simulate a cyber incident in a time-structured way while teams execute their real processes. The emphasis is on coordination, workflow execution, and organizational performance rather than on attacking or defending a live technical environment. These exercises validate the operational readiness of incident response and continuity mechanisms, including handovers between teams, decision-rights, task prioritisation, stakeholder management, and integration with external parties (e.g., service providers, CERTs, regulators). They are especially valuable for identifying “friction points” such as unclear responsibilities, delayed approvals, missing contact paths, and ineffective collaboration across organizational boundaries.

Technical drills and tests (operational validation exercises)

These exercises validate technical procedures and tool-supported response tasks. They focus on the concrete execution of detection, containment, eradication, and recovery steps. They often test whether tooling and operational prerequisites are available and functional (e.g., access rights, logging coverage, backup restoration, endpoint isolation, segmentation, incident ticket workflows). These drills can be narrow and frequent, forming the practical basis for more complex simulations.

Full-scale / integrated simulations (end-to-end exercises)

These exercises combine organizational coordination with technical response elements to simulate an incident as an end-to-end crisis. The aim is to test the system-of-systems: governance, operations, technical containment, communication, and recovery as an integrated capability. These formats tend to require substantial preparation and facilitation effort, but provide the highest realism and yield high-value findings about interdependencies and systemic gaps.

Core design principles aligned with BSI 100-4

A key contribution of the BSI approach is that it treats exercises as controlled projects with explicit design requirements. In the cyber domain, this means an exercise is defined not by “what happens,” but by what is **intentionally validated** and what evidence is generated to support improvement.

First, BSI-style exercise design starts with **clear objectives**. Cyber exercises should be derived from risk exposure and organizational priorities, such as improving incident response maturity, validating crisis governance, testing recovery time objectives, or rehearsing communication and reporting obligations. Objectives must be specific enough to be observed and evaluated, rather than broad aspirations like “be prepared.”

Second, the exercise scope must be defined carefully, including participating functions, external dependencies, and interfaces. In cyber crises, many weaknesses occur at boundaries ,between security and IT operations, between internal teams and suppliers, or between operational response

and executive decision-making. Therefore, a BSI-aligned approach prioritises the explicit modelling of roles, interfaces, and escalation paths.

Third, exercises should be built around scenarios that are realistic and relevant. A scenario is not merely a narrative; it is a structured sequence of stimuli and events that drive participant actions and enable observation. For cyber exercises, scenario realism requires alignment with threat models, typical attack paths, and organizational context (e.g., sector-specific impacts, regulatory reporting thresholds, service dependencies).

Fourth, BSI's logic implies that exercises require **controlled conditions**. This includes defining exercise artificialities (what is "assumed"), rules of engagement, and abort criteria. In cyber environments, these controls are essential to avoid operational risk and to maintain clarity about which observations are valid.

Documentation and artefacts for repeatability and auditability

A major strength of the BSI approach is the emphasis on documentation as a mechanism for quality, repeatability, and institutional learning. In cyber exercises, especially those intended for organizational benchmarking or national programmes, documentation also supports traceability and auditability.

A BSI-inspired cyber exercise documentation set typically includes:

- An **exercise plan** that anchors exercises over time and ensures coverage across business areas, processes, and maturity needs. This plan enables organizations to avoid ad-hoc selection and ensures that exercises match risk priorities.
- An **exercise fact sheet** summarising the key parameters (topic, format, target groups, scope, objectives, timeframe, constraints), supporting clarity for participants and stakeholders.
- An **exercise concept** describing the design in depth: objectives, scenario assumptions, participating roles, governance structure, facilitation approach, evaluation strategy, and logistics.
- An **inject plan / exercise script** specifying the detailed event flow and stimuli, including timing, delivery method, expected reactions, and facilitator guidance. This is essential for consistent execution and controlled escalation.
- Supporting templates for **invitations, briefings, participant guidance, observation protocols, feedback collection**, and structured reporting.

This artefact approach ensures that exercises can be repeated, improved, and compared across cycles. It also enables the organization to demonstrate that exercises were not only conducted, but designed to generate measurable learning and capability evidence.

Exercise type	Short description	Typical purpose	Relative effort
Test of technical precautionary measures	Technical verification of preventive or protective security measures	To assess whether technical safeguards are correctly implemented and functioning as intended	**1**
Functional test	Testing the functionality of selected technical or organizational measures	To verify whether specific functions, procedures, or controls work in practice	**2**
Plan review	Structured review of existing plans and documentation	To check completeness, consistency, and suitability of	**1**

		plans before practical execution	
Plan discussion	Discussion-based examination of plans with relevant stakeholders	To validate understanding, responsibilities, and procedural logic	**1–2**
Staff exercise	Exercise involving crisis staff or management teams	To train coordination, decision-making, and internal management processes	**1–2**
Command-post exercise	More advanced staff-based exercise with several roles and coordinated processes	To test command structures, communication flows, and leadership under pressure	**2–3**
Communication and alerting exercise	Focused exercise on notification, escalation, and communication procedures	To verify alarm chains, contact structures, and reporting mechanisms	**1**
Scenario simulation	Simulation of one or more scenarios in a structured exercise setting	To train response behaviour, cross-functional interaction, and scenario-based decision-making	**3**
Live or full-scale exercise	Realistic practical exercise under near-operational conditions	To validate the performance of people, processes, and technology in an integrated way	**4**

TABLE 3 - CYBER EXERCISE TYPES ACCORDING TO BSI 100-4

Table 3 summarises the exercise types distinguished in the BSI 100-4 framework. The classification is useful because it differentiates exercise formats not only by content and purpose, but also by their relative implementation effort. It therefore provides a practical reference point for structuring exercise portfolios and selecting appropriate formats depending on maturity, objectives, and available resources. Effort levels range from 1 (low) to 4 (high). The table is a cleaned and translated adaptation based on the BSI 100-4 exercise taxonomy.

Evaluation, learning, and continuous improvement

BSI 100-4 embeds exercises into a continuous improvement logic: an exercise is only valuable if it produces validated findings and drives action. In cyber exercises, evaluation should therefore be planned from the beginning, not added after execution.

A BSI-aligned evaluation approach typically includes:

- Defined **observation criteria** linked directly to objectives (e.g., timeliness of detection, escalation correctness, decision speed, communication quality, recovery effectiveness).
- Clear assignment of **observers and roles**, including who records observations, who moderates, and who owns the collection of evidence.
- A structured **debrief and after-action review** to consolidate findings and identify root causes rather than symptoms.
- A **remediation plan** that assigns ownership, prioritises corrective actions, sets deadlines, and identifies dependencies (e.g., process change, training need, tooling gap).
- **Re-testing** in subsequent exercises to verify that corrective actions improved capability, thereby turning exercises into a measurable improvement cycle.

Relevance for cyber exercise programmes and cyber range integration

Finally, BSI 100-4 provides a foundation that scales beyond a single organization. Its disciplined approach to objectives, documentation, and evaluation is compatible with sectoral and national cyber exercise programmes where comparability and quality assurance become strategic

requirements. It also aligns naturally with cyber range environments, because range-based exercises benefit from clear scenario design, inject control, observer methods, and post-exercise reporting.

In summary, BSI 100-4 supports a cyber exercise practice that is structured, repeatable, and improvement-driven. It enables organizations to develop cyber exercises as a core resilience mechanism: testing governance and coordination, validating technical response and recovery, and translating lessons learned into demonstrable capability enhancement over time.

1.3.1.4 *BSI 200-4 und Vorlage Übungskonzept*

The BSI Standard 200-4 (Business Continuity Management, BCM) gives you a clear, step-by-step way to create, maintain, and continuously upgrading a Business Continuity Management System (BCMS). It sees BCM as a management area that focuses on retaining or restoring time-sensitive services and processes when anything goes wrong, whether an IT outage, a cyber event, a supplier failure, or a problem with the facilities. The standard tells organizations to go from BCM governance and setup (policy, scope, roles, responsibilities, and management commitment) to a systematic analysis of critical activities and dependencies, usually using a Business Impact Analysis (BIA) to set priorities and recovery goals. Based on this, the organization makes and follows through on plans and documents (like response and recovery procedures, communication and escalation paths, and resource requirements) that are needed for business continuity management (BCM) to be a part of everyday business. One crucial principle is that just having continuity plans doesn't guarantee they're excellent; they need to be checked and improved through systematic verification, such exercises and testing, with evidence-based evaluation leading to repeated improvements in capabilities and documentation.

The BSI "Übungskonzept" template (Exercise Concept template) puts these standards into practice by giving all parts of an organization a consistent, auditable structure for a BCM exercise program. The template helps businesses lay down the goals, scope, assumptions, and purpose of the exercise program in a way that is always the same. It also pushes them to create a clear list of exercise types that meet with BCM goals and the organization's level of maturity. These could include plan/document reviews, alarm and communication drills, crisis management or staff exercises, and technical recovery or restoration testing. It also sets up the governance parts needed for reliable execution, such as roles and responsibilities for planning, control, observation/evaluation, and approvals, as well as organizational and technical requirements (participants, connections to outside providers, communication channels, escalation logic, and logistical plans). The template is vital since it helps the cycle of constant improvement. You should keep track of the results of exercises in a consistent fashion, search for gaps in capacity and procedure, turn them into assigned corrective actions with owners and dates, and then check them again by doing follow-up tests or re-exercising. The BSI method doesn't regard exercises as things that happen once. Instead, they are a way to show and strengthen an organization's resilience over time.

BSI 200-4 and the Exercise Concept (Übungskonzept) template can be used as a governance and quality backbone for a national cyber exercise program. This is especially true when the national goal is to make improvements in comparability, auditability, and measurable resilience across sectors. The BSI approach can be seen as a common framework for continuity and crisis response at the national level. For example, the BCMS scope definition, BIA-driven criticality logic, recovery targets, and continuity strategies can all be turned into national baseline requirements, such as minimum continuity objectives for essential services, sector-specific dependency modeling, and standard expectations for crisis roles and escalation paths. This sets a "minimum standard" for what each participating group must be able to show in exercises, but it also lets each group construct scenarios that fit its own situation.

The Exercise Concept (Übungskonzept) helps national programs by making it possible for various organizations to plan and document exercises in the same way. A national authority (or program

office) can force participants to use the same framework for their exercise concepts. This makes it possible to compare and review workouts from cycle to cycle. You can map different types of exercises to a national progression model. For example, you can move from document reviews and communication drills to cross-organization staff exercises and technical recovery tests. This lets national planners manage participation based on maturity and give more difficult operational exercises to groups that have already shown baseline readiness. Also, the template's descriptions of roles and responsibilities can be utilized to make national role profiles more consistent (for example, crisis manager, technical lead, communications lead, liaison officer). This will make it easier for different organizations to work together during multi-entity exercises and genuine occurrences.

The BSI logic can make national programs stronger by making it clear how they might be improved. If the national program requires that exercise results be turned into remedial actions with owners and deadlines, and that progress be checked by follow-up tests, then the program can produce measurable national-level results, not just "exercise completion." A nationwide AAR/IP monitoring system can make this happen. Each organization keeps track of its own action plans, and the software combines anonymous (or secret, if needed) data into ongoing systemic gaps, sector trends, and cross-cutting dependency issues. Over time, this lets the government of a country move from learning on the fly to managing resilience based on evidence. In this new system, recurring weaknesses (like third-party dependencies, decision-making delays, communication breakdowns, and restoration bottlenecks) are given priority and tested again in the next exercise cycle.

Finally, BSI 200-4 is very helpful for national cyber exercise programs since it links cyber incident response to service continuity and the effects on society. National exercises can do more than just test technical detection and containment. They can also test whether essential services can be kept going or brought back to life within certain time limits, whether crisis governance works under pressure, and whether inter-organizational coordination helps continuity goals. The BSI framework can add to cyber-specific exercise methods (like incident-handling or red-team/blue-team formats) by giving them a shared measurement layer that focuses on continuity and supports national KPIs, sector benchmarking, and ongoing improvement across the national ecosystem.

1.4 EU Strategies and initiatives

In this chapter, I will list and discuss the importance of the following key regulations and directives such as

- the Network and Information Security Directive (known as NIS2)[50] ,
- the EU Cybersecurity Act[51],
- the Cyber Resilience Act[52], and Critical Entities Resilience Directive (CER)[53]

[50] European Parliament and Council of the European Union. (2022). *Directive (EU) 2022/2555 of 14 December 2022 on measures for a high common level of cybersecurity across the Union (NIS2 Directive). Official Journal of the European Union, L 333*, 80–152.
Official NIS2 text on EUR-Lex

[51] European Parliament and Council of the European Union. (2019). *Regulation (EU) 2019/881 of 17 April 2019 on ENISA and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act). Official Journal of the European Union, L 151*, 15–69.
Official Cybersecurity Act on EUR-Lex

[52] European Parliament and Council of the European Union. (2024). *Regulation (EU) 2024/2847 of 23 October 2024 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.*
Official Cyber Resilience Act on EUR-Lex

[53] European Parliament and Council of the European Union. (2022). *Directive (EU) 2022/2557 of 14 December 2022 on the resilience of critical entities and repealing Council Directive 2008/114/EC*

- the Digital Operational Resilience Act[54],
- the General Data Protection Regulation (GDPR)[55], and
- the EU Cyber Blueprint[56].

Network and Information Systems Directive (NIS und NIS2)

The **NIS2 Directive (Directive (EU) 2022/2555)**[57] represents one of the central regulatory pillars of the European Union's cybersecurity framework. It replaced the original NIS Directive (Directive (EU) 2016/1148) and substantially expanded both the scope of regulated entities and the governance expectations associated with cybersecurity risk management. Its overarching objective is to achieve a **high common level of cybersecurity across the Union** by strengthening national cybersecurity capabilities, improving cooperation between Member States, and introducing more consistent cybersecurity risk-management, incident-reporting, supervision, and enforcement requirements.

NIS2 establishes obligations at both **national and organizational level**. Member States are required to maintain national cybersecurity strategies and designate competent authorities, cyber crisis-management authorities, single points of contact, and Computer Security Incident Response Teams (CSIRTs). At European level, the Directive reinforces cooperation through the **Cooperation Group**, the **CSIRTs Network**[58], and the **European Cyber Crisis Liaison Organisation Network (EU-CyCLONe)**[59]. EU-CyCLONe is particularly relevant for large-scale incidents because it supports the coordinated management of cybersecurity incidents and crises at operational level and facilitates the exchange of information between Member States and EU institutions.

Compared with the original NIS Directive, NIS2 significantly broadens the range of entities subject to cybersecurity obligations. It covers **essential and important entities** across sectors including energy, transport, banking and financial-market infrastructure, health, digital infrastructure, public administration, space, postal and courier services, waste management, manufacturing of critical products, digital providers, and research. This broader sectoral scope reflects the growing recognition that cyber resilience depends on interconnected ecosystems rather than on the protection of individual organizations in isolation.

(Critical Entities Resilience Directive). Official Journal of the European Union, L 333, 164–198. Official CER Directive on EUR-Lex

[54] European Parliament and Council of the European Union. (2022). *Regulation (EU) 2022/2554 of 14 December 2022 on digital operational resilience for the financial sector (Digital Operational Resilience Act). Official Journal of the European Union, L 333*, 1–79. Official DORA text on EUR-Lex

[55] European Parliament and Council of the European Union. (2016). *Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, L 119*, 1–88. Official GDPR text on EUR-Lex

[56] Council of the European Union. (2025). *Council Recommendation of 6 June 2025 on an EU blueprint for cyber crisis management (C/2025/3445). Official Journal of the European Union, C/2025/3445*. Official EU Cyber Blueprint on EUR-Lex

[57] European Parliament and Council of the European Union. (2022). *Directive (EU) 2022/2555 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 (NIS2 Directive). Official Journal of the European Union, L 333*, 80–152. Official NIS2 Directive on EUR-Lex

[58] European Union Agency for Cybersecurity. (n.d.). *CSIRTs Network*. ENISA. Official CSIRTs Network page

[59] European Union Agency for Cybersecurity. (n.d.). *EU CyCLONe*. ENISA. Official EU-CyCLONe page

A central element of the Directive is the requirement for appropriate and proportionate **cybersecurity risk-management measures**. These measures address areas such as incident handling, business continuity and disaster recovery, crisis management, supply-chain security, vulnerability management, policies for assessing the effectiveness of cybersecurity measures, cyber hygiene and training, access control, asset management, and the use of appropriate cryptographic solutions. Management bodies are also required to approve cybersecurity risk-management measures and oversee their implementation, strengthening the link between technical cybersecurity and organizational accountability.

This shift is particularly important from the perspective of cyber exercises. NIS2 does not establish one mandatory European cyber exercise format for all regulated entities. However, several of its requirements concern capabilities whose effectiveness cannot be demonstrated reliably through policy documents alone. Incident handling, escalation, continuity arrangements, crisis-management responsibilities, supply-chain coordination, management decision-making, and recovery procedures can all be **tested and validated through cyber exercises**. Exercises can therefore provide practical evidence of whether formally established cybersecurity arrangements function effectively under realistic conditions.

For example, an exercise may examine whether a significant incident is identified and escalated within the required organizational structure, whether responsible managers receive sufficient information to make decisions, whether continuity and recovery procedures can be activated, whether dependencies on external ICT providers are understood, and whether communication with competent authorities can be performed within regulatory time constraints. In this sense, NIS2 contributes to a broader transition from cybersecurity as a predominantly technical-control problem towards **governed and demonstrable operational resilience**.

Incident reporting provides another important connection to exercise design. NIS2 introduces a staged reporting mechanism for significant incidents, including an **early warning within 24 hours**, an incident notification generally within **72 hours**, and a final report normally within **one month**. These time-dependent obligations can be incorporated directly into tabletop, functional, or cross-organizational exercises. A realistic scenario can therefore test not only whether an incident is technically detected, but whether the organization can generate an adequate situational picture, determine regulatory relevance, coordinate internal approvals, and communicate with the responsible authority within the required timeframe.

NIS2 is also relevant at the **national exercise ecosystem level**. The Directive requires national strategies, competent authorities, CSIRTs, crisis-management structures, and cross-border cooperation mechanisms. These structures represent many of the same actors that must cooperate during large-scale or national cyber exercises. Consequently, national exercises can serve as controlled environments for testing whether institutional roles, information-sharing arrangements, escalation mechanisms, and cross-sector coordination structures operate as intended.

From the perspective of this dissertation, NIS2 therefore acts as an important **regulatory convergence driver**. It does not harmonize cyber exercise scenarios or technologies, but it increasingly harmonizes the capabilities that European organizations and national authorities are expected to possess and demonstrate. These include governance and accountability, risk management, incident preparedness, continuity, supply-chain resilience, information sharing, reporting, documentation, and continuous improvement. Such requirements directly support several dimensions of the European Convergence Framework developed in Chapter 1.5, particularly **governance and accountability, lifecycle orientation, evaluation and measurement, auditability and evidence, dependency and ecosystem management, and continuous improvement**.

Accordingly, NIS2 should not be interpreted as creating a uniform European cyber exercise methodology. Its significance for cyber exercising lies instead in establishing a common regulatory baseline against which preparedness and resilience increasingly need to be demonstrated. Cyber

exercises provide one practical mechanism for translating these regulatory expectations into observable behaviour and evidence, thereby connecting formal cybersecurity governance with operational readiness.

The implementation of NIS2 at Member State level is examined further in **Part II**, where Austria serves as the principal national reference case and the European requirements are analysed in relation to national legislation, competent authorities, coordination structures, and the proposed National Cyber Exercise Governance Model.

EU Cybersecurity Act

The EU Cybersecurity Act is Regulation (EU) 2019/881, which was passed on April 17, 2019, and published in the Official Journal in June 2019. It has two main parts: (1) it gives ENISA (the European Union Agency for Cybersecurity) a permanent mandate and sets out its duties and governance; and (2) it sets up a framework for EU-wide European cybersecurity certification schemes for ICT products, services, and processes. This is meant to make sure that claims of security are the same across the EU and to cut down on fragmentation in the internal market.

The Act makes ENISA a key player in EU-level collaboration and fostering cybersecurity skills from a policy and governance point of view. The Act is most recognized for certification, but its ENISA pillar is directly related to cyber exercises since it makes ENISA's role in helping Member States and EU institutions get ready, work together, and come up with shared methods official. In practice, this relates to exercises through ENISA's long-standing management of large-scale, cross-border cyber crisis management exercises, such as the Cyber Europe series, which simulate escalating cyber incidents and test coordination mechanisms and decision-making at the European level.

The Cybersecurity Act can be utilized as a helpful reference in three different ways when designing a national (or multinational) cyber exercise program. First, it backs up the idea of interoperability and comparability: a national program can make its exercise governance fit with the Act's EU-level cooperation logic by using the same language, having the same role expectations, and setting cross-entity coordination goals that make it easier for people to take part in EU-level exercises (like Cyber Europe) and learn from them to improve national preparedness. Second, the Act's certification pillar can indirectly affect the scope of an exercise by getting organizations to use the cybersecurity features of certified ICT products, services, and processes as assurance baselines. This way, exercises can check to see if operational practices (configuration, monitoring, incident response, continuity) really do meet the expected level of assurance in real decision-making situations, such as when there are suppliers and dependencies. Third, the Act makes the case for a continuous-improvement loop at the ecosystem level even stronger. Exercises become a formal way to find systemic coordination gaps and send them to national and EU cooperation institutions where lessons, good practices, and capability requirements are put together.

Critical Entities Resilience Directive (CER) [60]

The Critical Entities Resilience (CER) Directive, which was enacted on December 14, 2022, is Directive (EU) 2022/2557[61]. It sets out the EU's legal framework for making sure that critical entities

[60] European Union, "DIRECTIVE (EU) 2022/2557 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 December 2022 on the resilience of critical entities and repealing Council Directive 2008/114/EC," Official Journal of the European Union, no. 333, 2022.

[61] European Parliament and Council of the European Union. (2022). *Directive (EU) 2022/2557 of 14 December 2022 on the resilience of critical entities and repealing Council Directive 2008/114/EC (Critical Entities Resilience Directive). Official Journal of the European Union, L 333*, 164–198. Official CER Directive on EUR-Lex

that provide important services are better able to deal with a wide range of risks. It went into effect on January 16, 2023, and by October 17, 2024, all Member States had to make it part of their own laws. CER looks at all types of hazards, including natural disasters, accidents, sabotage, insider threats, terrorism, public health emergencies, and hybrid disruptions. It also broadens the EU's resilience policy beyond the previous European Critical Infrastructure approach by encompassing more industries. The Directive covers 11 areas, such as energy, transportation, banking, financial market infrastructure, health, drinking water, wastewater, digital infrastructure, public administration, space, and food. This shows that the EU believes that society can only continue to exist if there are connections between different sectors and countries.

In terms of methodology, CER sets up a governance chain that connects national risk understanding to entity-level resilience measures and oversight. Member States are required to perform risk assessments and utilize the results to facilitate the identification of essential entities and to inform national resilience priorities. In turn, critical entities are expected to take appropriate technological, security, and organizational steps to stop incidents, deal with interruptions, and get important services back up and running. The Directive also sets up a stronger structure for oversight and coordination. This includes competent authorities and processes that let Member States (and the EU) check to see if resilience measures are being put into practice and not only written down.

Even if the Directive isn't just about cyber threats, CER is directly and operationally linked to cyber exercises. Cyber incidents are a big reason why services are disrupted in most CER sectors. Exercises are a safe way to see if service continuity and crisis governance work under real-world cyber pressure (for example, ransomware that affects hospital operations, compromised OT monitoring that affects drinking water treatment, and supplier outages that affect emergency communications). Second, CER makes it clear that the EU has a mandate to build capabilities through guidance, methodologies, and cross-border training activities and exercises. The Commission's job is to support Member States' efforts by coming up with best practices and organizing cross-border exercises to test the resilience of important entities. This clause makes exercises a legally supported way to test resilience measures, not just a "good practice" that people can choose to do.

CER can be used as a design anchor for national cyber exercise programs. This changes the focus of the evaluation from "technical response success" to the impact on society and services. In reality, a national program can get exercise goals from CER obligations, such as being able to keep important services running, making decisions when there is uncertainty, communicating during a crisis, managing dependencies, and setting recovery priorities. It can also run maturity-based exercise cycles that go from discussion-based validation to operational coordination between multiple entities. Importantly, CER makes the case for an improvement loop stronger: results from exercises may be linked to resilience needs and tracked as corrective actions with responsible owners and due dates. This makes it possible to show progress over the course of annual or multi-annual exercise cycles.

Lastly, CER helps raise awareness in Europe by making resilience a regular topic of discussion for leaders and governments in all Member States. It gives the EU a common language for what "resilience of essential services" means, sets deadlines and expectations for supervisors that bring resilience to the attention of boards and ministers, and explicitly encourages cross-border cooperation. This is an important way to raise awareness because cross-border exercises show interdependencies that are not visible in single-organization drills. The Commission has also given support and advice on how to put things into action, including more recent guidelines to help nations consistently implement CER. This has helped everyone understand each other better and sped up the process of everyone in the EU adopting the same rules.

Digital Operational Resilience Act (DORA)

The Digital Operational Resilience Act (DORA) is Regulation (EU) 2022/2554[62], which was passed in 2022 as part of the EU's larger digital-finance package. Its goal is to make it easier to manage ICT and cyber risk in the financial system across all sectors. DORA wants to make sure that financial institutions can handle, respond to, and recover from ICT-related problems, such as cyberattacks. It does this by bringing together laws that were spread out across different sectors. Specifically, it establishes duties in five interconnected domains: (1) governance and controls for ICT risk management (including policies, roles, asset and dependency management, protection, detection, response, and recovery measures); (2) incident management and reporting related to ICT; (3) testing for digital operational resilience; (4) third-party risk management for ICT (including cloud and other ICT providers); and (5) arrangements for sharing information. The Regulation makes it clear that these steps are meant to make people more aware of ICT risks and make the EU financial sector less vulnerable as a whole.

DORA defines **Threat-Led Penetration Testing** (TLPT) as the highest level of resilience testing. It falls under the "advanced testing" rules (specifically Articles 26–27) and is meant to mimic real, intelligence-led attacks on the important or critical functions of a financial entity and the systems and services that support those functions (including those provided by ICT third-party providers). TLPT is not a standard "point-in-time" penetration test; it is a controlled, end-to-end exercise that tests whether an organization can find, stop, and recover from advanced enemy behavior while keeping production stable and reducing unexpected effects. The legal framework also makes it easier for supervisors to work together and recognize TLPT results across Member States. This is vital for cross-border groups and harmonized assurance.

Methodologically, TLPT under DORA is closely connected with the EU's existing TIBER-EU[63] method to threat intelligence-based ethical red teaming. TIBER-EU sets up an end-to-end system for intelligence-led testing in live production systems. This includes governance, defining important functions, developing threat intelligence, controlled red teaming, and a formal "replay" and remediation phase. At the DORA level, TLPT materials make it clear that organizations can utilize TIBER-EU or national TIBER implementations as long as they meet DORA's requirements. This connects regulatory testing to a mature European testing lifecycle and shared templates.

In terms of connection to cyber exercises, DORA effectively turns cyber resilience testing into a regulated, auditable form of exercise programming for the financial sector. A DORA-aligned exercise portfolio can be structured as a maturity ladder: from basic tests (e.g., vulnerability assessments, scenario-based tabletop exercises for ICT incidents, backup/restore validation) to integrated operational tests, and finally to TLPT as an intelligence-led “full-spectrum” exercise that stresses governance, SOC operations, crisis management, third-party coordination, and service continuity simultaneously. TLPT results can feed directly into an After-Action / Improvement Plan cycle: findings are translated into remedial actions with accountable owners and timelines, and follow-up tests validate that remediation reduced the originally observed exposure ,creating a measurable resilience loop that is suitable for both supervision and internal risk governance.

Finally, DORA and TLPT support cybersecurity awareness in Europe by institutionalising ICT risk as a board-level and supervisory priority and by standardising “what good looks like” across the single market. Because DORA applies consistently across many financial entity types and explicitly addresses third-party dependencies, it helps shift awareness from isolated organizational security

[62] European Parliament and Council of the European Union. (2022). *Regulation (EU) 2022/2554 of 14 December 2022 on digital operational resilience for the financial sector (Digital Operational Resilience Act). Official Journal of the European Union, L 333*, 1–79.
Official DORA text on EUR-Lex

[63] European Central Bank. (2025). *TIBER-EU framework*.
https://www.ecb.europa.eu/pub/pdf/other/ecb.tiber_eu_framework_2025~b32eff9a10.en.pdf
Official TIBER-EU Framework 2025

to ecosystem resilience, including cloud concentration risk, cross-entity contagion, and operational interdependence. TLPT amplifies this awareness further by making “realistic adversary thinking” and evidence-based learning an expected practice, while EU-level alignment with TIBER-EU encourages common language, comparable outcomes, and cross-border learning across Member States and supervisory communities.[64]

Data Protection Law (GDPR)

In data protection law, the principle of prohibition applies: the use of data is, in principle, not permitted unless its permissibility follows from the law. The permissibility of each intended use of data must therefore be examined. The examination of data protection law permissibility for an intended data processing operation is to be carried out according to the following scheme, which derives from §§ 6 para. 1 and 7 paras. 1 and 3 DSG:

1. Examination of the authorisation of the controller
2. Consideration of the legitimate secrecy interests of the data subjects
3. Review of the necessary extent and the least intrusive means
4. Examination of compliance with the general principles of § 6 DSG

Examination of the authorisation of the controller

§ 7 para. 1 DSG stipulates that data may only be processed within the framework of the statutory competences or legal powers of the respective controller. The requirement of statutory competence refers to controllers in the public sector and is fulfilled if there is a sufficiently determined authorisation in substantive law for the use of the data, or if the right to use the respective data results from the statutory remit of the authority. This gives effect to the principle of legality described above in Art. 18 para. 1 B-VG.

The requirement of legal power relates primarily, but not exclusively, to controllers in the private sector. Legal power may arise, for example, from a business licence or concession, but also from general rights (property rights, rights of domicile, etc.) and duties of the controller.

Consideration of the legitimate secrecy interests of the data subjects

The processing of data is only permissible insofar as it does not violate the legitimate secrecy interests of the data subject. This is laid down in § 7 para. 1 DSG and is specified, for non-sensitive data, in § 8 DSG and, for sensitive data, in § 9 DSG. The grounds of permissibility in § 8 DSG are not exhaustive; besides fulfilling one of the explicitly listed grounds of permissibility, the permissibility of processing non-sensitive data under § 8 para. 1 item 4 DSG may also result from a balancing of interests, where legitimate interests of the controller or of a third party outweigh the legitimate secrecy interests of the data subject.

The following grounds of permissibility in § 8 DSG are particularly relevant in the present context:

Explicit statutory authorisation or obligation to use the data

Where there is an explicit statutory authorisation or obligation to use the data, legitimate secrecy interests are deemed not to be violated pursuant to § 8 para. 1 item 1 DSG. High requirements are placed on the degree of determination of such statutory authorisations or obligations, so that the occasion and purpose, types of data, group of potentially affected persons etc. must be precisely defined in the respective provision granting such powers.

[64] European Union, "REGULATION (EU) 2022/2554 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC)," Official Journal of the European Union, no. 333, 27 12 2022.

This ground of permissibility applies to controllers both in the public and in the private sector. Since the fundamental-rights requirement in § 1 para. 2 DSG, on which this ground of permissibility is based, only applies to interferences by state authorities, less stringent requirements exist for the degree of determination of statutory authorisations or obligations addressed to private controllers.

Overriding legitimate interests of the controller or of a third party

Legitimate secrecy interests are also deemed not to be violated, in the case of non-sensitive data, where overriding legitimate interests of the controller or of a third party require the use of the data (§ 8 para. 1 item 4 DSG). This is particularly the case, inter alia, where the use of the data is an essential prerequisite for a controller in the public sector to perform a task conferred on it by law (§ 8 para. 3 item 1 DSG), where it occurs in fulfilment of the obligation to provide administrative assistance, or where it is required to fulfil a contractual obligation between the controller and the data subject (§ 8 para. 3 item 1 DSG).

Consent of the data subject

If the data subject has consented to the use of his or her data, legitimate secrecy interests are deemed not to be violated under § 8 para. 1 item 2 DSG. If, for example, a company is affected by a security incident and therefore contacts a situation centre on its own initiative, this can be regarded as implied consent to the use of the data thus transmitted. However, this only covers data relating to the company, not data of, for instance, customers or business partners, because in relation to such persons the company is the controller and not itself the data subject. It must also be borne in mind that the data subject may withdraw his or her consent at any time, which "renders the further use of the data impermissible" (§ 8 para. 1 item 2 DSG).

Published data and indirectly personal data

According to § 9 item 1 DSG, legitimate secrecy interests are not violated when using sensitive data if the data subject has obviously made the data public himself or herself. For non-sensitive data, § 8 para. 2 DSG provides that, when using lawfully published data, legitimate secrecy interests are generally deemed not to be violated. When lawfully published data and only indirectly personal data are used, legitimate secrecy interests are therefore deemed not to be violated.

Review of the necessary extent and the least intrusive means

According to § 7 para. 3 DSG, the permissibility of data use also requires that it occur only to the necessary extent and by using the least intrusive means available. This is the statutory elaboration of the fundamental-rights principle of proportionality and means that the use of data is lawful only if it is suitable and necessary as a means to achieve a specific objective. If several means meet these conditions, the least intrusive one – that is, the one with the lowest intensity of interference – must be chosen. The requirement to use the least intrusive means does not go so far as to render the use of electronic data processing systems impermissible merely because the given purpose could also be achieved by other, in particular non-electronic, means. It is therefore not necessary to examine whether the respective purpose could also be achieved without the use of electronic data processing systems.

Examination of compliance with the general principles of § 6 DSG

Pursuant to § 7 para. 3 DSG, the permissibility of data use further requires compliance with the principles set out in § 6. In transposition of essential general principles of the European data protection tradition, § 6 DSG lays down further conditions that must be observed when using personal data:

- **Fairness and lawfulness:** Data may only be used in good faith and in a lawful manner (§ 6 para. 1 item 1 DSG). According to the explanatory materials, this is only fulfilled "if the data subject is not misled or left in the dark about the circumstances of data use and the existence and enforceability of his or her rights."

- **Strict purpose limitation:** Data may only be collected for specified, explicit and legitimate purposes and may not be further used in a manner incompatible with those purposes (§ 6 para. 1 item 2 DSG).
- **Limitation of data volume (principle of necessity):** Data may only be used insofar as they are essential for the purpose of the data application and may not go beyond that purpose (§ 6 para. 1 item 3 DSG).
- **Accuracy and up-to-dateness:** Data may only be used in such a way that, in view of the purpose of use, they are factually correct in outcome and, where necessary, kept up to date (§ 6 para. 1 item 4 DSG).
- **Temporal limitation:** Data may only be stored in personal form for as long as is necessary to achieve the purposes for which they were collected; longer storage periods may arise from special statutory provisions, in particular those relating to archiving (§ 6 para. 1 item 5 DSG).

Criminal-law-relevant data

Data relevant under criminal law are considered particularly worthy of protection, which is why stricter rules apply to their use. In connection with situation pictures, such data can arise in security incidents with a criminal background (computer and internet criminal law), especially since even mere suspicion of the commission of criminal offences is to be classified as data relevant under criminal law. In the case of such data, legitimate secrecy interests of the data subject are only not violated if

- there is an explicit statutory authorisation or obligation to use such data, or
- the use of such data is an essential prerequisite for controllers in the public sector to perform a task conferred on them by law, or
- the permissibility of using such data otherwise results from statutory duties of care or from other legitimate interests of the controller that outweigh the legitimate secrecy interests of the data subject, and the manner in which the data application is carried out, or
- the disclosure of data is for the purpose of filing a report with an authority competent to prosecute the reported criminal acts (or omissions).

Reform of EU Data Protection Law

Despite the applicability of the Data Protection Convention and, in particular, the Data Protection Directive (DSRL), there is no uniform data protection law in the EU, since the DSRL, like any directive, had to be transposed into national law in each Member State, and these national transpositions differ in detail. The factual distortions of competition resulting from the different levels of data protection in the EU Member States have not been sufficiently eliminated by the DSRL. Among other reasons, the European Commission therefore presented, on 25 January 2012, the reform of EU data protection law with the aim of making Europe fit for the digital age.

As part of the strategy to realise the Digital Single Market and the European Security Agenda, the reform of EU data protection law comprises the proposal for a Regulation of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation, GDPR) and the proposal for a Directive on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties and on the free movement of such data. The GDPR was adopted by the European Parliament on 14 April 2016. It was published in the Official Journal of the European Union on 4 May 2016 and enters into force on 25 May 2016. It is applicable from 25 May 2018, meaning that from 25 May 2018 the Austrian Data Protection Act is no longer applicable, but instead the EU GDPR applies directly.

EU Cyber Blueprint

The EU Cyber Blueprint is the EU's non-binding, policy-level framework for coordinated cyber crisis management in the event of a large-scale, cross-border cybersecurity incident or cyber crisis. It is intended to make EU-level crisis handling more predictable and interoperable by clarifying who does what, when, and how information and decisions should flow across technical, operational, and political levels. In June 2025, the Council adopted an updated Cyber Blueprint that builds on and supersedes the earlier 2017 Commission "Blueprint" recommendation, reflecting lessons learned since 2017 and aligning with newer EU legislation and structures (notably those introduced or strengthened under NIS2).

Methodologically, the Cyber Blueprint is structured around a crisis lifecycle and defines coordination activities across the phases of preparedness, detection, response, public communication, recovery, and lessons learned. It maps the relevant EU-level actors and networks to these phases and explains how they should cooperate to achieve shared situational awareness and coordinated response measures. A central design principle is the separation ,but tight coupling ,of technical coordination (incident analysis, mitigation, technical reporting) with operational and political coordination (national coherence across sectors, EU-level coordination, and potential activation of EU crisis mechanisms)

One important thing the Blueprint does is make it clear that the current EU cyber crisis structures and mechanisms work together as a "system of systems." It expands on the work of the CSIRTs Network and the EU-CyCLONe network (European cyber crisis liaison organization network) at the technical-operational level. It also shows how ENISA can help with coordination and information aggregation. It connects cyber crisis management to bigger EU crisis coordination plans, such the Integrated Political Crisis Response (IPCR), so that cyber catastrophes that turn into bigger problems in society may be handled using tools from all levels of government. The Blueprint also stresses the importance of clear communication with the public and the necessity to coordinate messaging and information sharing to avoid conflicting stories during cross-border emergencies.

There is a clear and fundamental link to cyber exercises. The new Blueprint is based on what was learned from Union-level exercises and is expected to be evaluated and improved through a series of EU cyber exercises. In practice, this means that the Blueprint can be used as a guide for planning and judging strategic and operational coordination exercises. For example, it can be used to practice how CSIRTs Network reports are sent to EU-CyCLONe, how EU-CyCLONe briefs the Council, how ENISA-supported situational awareness products are put together, and when political crisis mechanisms are used. The Blueprint even says that civilian-military collaboration should be put into practice in the right situations. This makes it possible to construct joint exercises that test how civilian cyber crisis actors and military cyber stakeholders at the EU level can work together.

Finally, the EU Cyber Blueprint supports cyber resilience awareness in Europe by institutionalising a common European understanding of cyber crises as cross-border, multi-sector events that require coordinated action beyond purely technical incident response. By standardising roles, escalation logic, and coordination products, it increases the "readiness literacy" of national authorities and EU bodies, and it strengthens a shared culture of learning by embedding recovery and lessons learned as lifecycle phases rather than optional follow-ups. This creates a clearer governance pathway for turning exercise and incident lessons into improvements in procedures, cooperation arrangements, and preparedness across Member States.

Overview and Summary

Chapter 1.4 demonstrates that the European Union is increasingly shaping a common regulatory and policy environment for cybersecurity, resilience, data protection and crisis preparedness. Although the individual instruments examined in this chapter have different legal purposes and

sectoral scopes, they collectively influence the conditions under which organizations and national authorities are expected to prepare for, respond to and recover from cyber incidents. NIS2 strengthens cybersecurity governance, risk-management and incident-response obligations across a broad range of essential and important entities. The Cybersecurity Act reinforces the institutional and certification framework of the European cybersecurity ecosystem. CER extends resilience expectations beyond cybersecurity alone by focusing on the continuity and robustness of critical entities and essential services. DORA introduces particularly explicit requirements for digital operational resilience and testing in the financial sector, while GDPR establishes important boundaries for the lawful processing, handling and protection of personal data. The EU AI Act adds a further governance dimension by introducing requirements concerning responsible AI use, human oversight, accountability and risk management. Finally, the EU Cyber Blueprint strengthens the European crisis-management perspective by defining how large-scale cyber incidents should be coordinated across Member States and EU-level structures.

Taken together, these developments show that European cyber resilience is no longer shaped by isolated technical security requirements. The emerging regulatory environment increasingly combines governance, accountability, continuity, dependency management, testing, information sharing, crisis coordination, documentation and evidence. This is directly relevant to cyber exercises because exercises provide one of the practical mechanisms through which such capabilities can be tested under controlled but realistic conditions. Even where a regulation does not prescribe a specific cyber exercise format, it may still create expectations that can be translated into exercise objectives. For example, an exercise can test whether incident escalation works in practice, whether reporting responsibilities are understood, whether third-party dependencies are managed effectively, whether continuity arrangements can be activated, whether decision-making remains effective under time pressure, or whether sensitive data is handled appropriately during a crisis.

The analysis therefore shows that EU regulation and policy increasingly influence not only whether organizations should be prepared, but also what credible preparedness should look like. This has implications for exercise design, because objectives, scenarios, injects, evidence requirements and evaluation criteria can increasingly be linked to regulatory and governance expectations. At the same time, the chapter also makes clear that the European Union has not established one single mandatory cyber exercise model. Different sectors, organizations and Member States continue to operate under different conditions and use different testing and exercise formats. The convergence lies instead in the underlying expectations: cyber preparedness should be governed, documented, evidence-oriented, resilience-focused and capable of supporting continuous improvement.

Chapter 1.4 therefore provides the regulatory and policy foundation for the European Convergence Framework developed in the following section. The subsequent analysis builds on the observation that international exercise methodologies and European regulatory developments increasingly reinforce similar principles. These recurring principles can then be synthesised into a common baseline that allows heterogeneous cyber exercise practices to become more comparable and auditable without requiring uniform implementation.

1.5 European Convergence Framework (C1)

The preceding sections demonstrate that the international and European cyber exercise landscape is characterised by an apparent tension. On the one hand, cyber exercises remain heterogeneous in terminology, purpose, governance, technical depth, documentation, evaluation, and institutional setting. A strategic crisis-management tabletop exercise, a national multi-sector coordination exercise, a live-fire technical exercise, and a threat-led resilience test may all contribute to cyber resilience while following substantially different methodologies and producing different types of evidence. On the other hand, the analysis of established exercise methodologies, recurring international exercise programmes, and European regulatory and policy instruments reveals a

growing set of common expectations concerning how cyber exercises should be designed, governed, documented, evaluated, and translated into improvement.

This section synthesises these findings into the first main contribution of the dissertation: the **European Convergence Framework for Cyber Exercises**.

The framework does not propose a single European exercise methodology and does not assume that all cyber exercises should become identical. Such an approach would neither be realistic nor desirable because exercise objectives, participating sectors, organizational maturity, confidentiality requirements, threat environments, and technical capabilities differ substantially. Instead, the framework identifies the **dimensions in which convergence is already emerging despite this contextual diversity**.

The central proposition is that European convergence is taking place less at the level of individual scenarios or exercise formats and more at the level of the **underlying governance and quality principles**. Different organizations may conduct different exercises, but increasingly they are expected to demonstrate why the exercise was conducted, what was tested, how responsibilities were assigned, what evidence was collected, what weaknesses were identified, and what was improved afterwards.

The framework therefore distinguishes between **harmonisation of exercises** and **convergence of exercise governance**. Full harmonisation would imply largely uniform exercise formats, scenarios, procedures, or evaluation systems. The evidence analysed in this dissertation does not support such a conclusion. Convergence, by contrast, means that heterogeneous exercise systems increasingly adopt or are subject to similar minimum expectations for quality, documentation, accountability, evidence generation, and continuous improvement.

From Fragmentation to Convergence

The starting point of the framework is the fragmentation identified throughout Part I. International and European cyber exercise practices differ along several dimensions. Different terminology is used for exercises, simulations, stress tests, cyber ranges, red teaming, resilience testing, and threat-led testing. Exercise formats range from executive discussion-based activities to functional simulations and technically intensive live-fire environments. Governance arrangements also vary: some exercises are conducted internally by organizations, others are coordinated by sectoral bodies, national authorities, EU institutions, defence organizations, research institutions, or international alliances.

The interaction between national strategic development and international cybersecurity expectations is illustrated by Magyar et al. (2025), whose comparative analysis of Hungarian cybersecurity strategies identifies increasing engagement with EU and NATO objectives alongside continued national institutional and strategic development. This example supports the distinction between convergence and uniformity: shared international principles may influence national cybersecurity governance while implementation remains shaped by domestic priorities, institutional structures, and strategic requirements.[65]

International convergence in cyberspace remains constrained by differing strategic interests and competing approaches to cyber governance. Molnár (2022) illustrates these divergences through the cyber-diplomatic positions of China, the United States, and Russia, demonstrating that cyber

[65] Magyar, S., Bányász, P., Bányász-Váczi, K. B., & Pál, A. (2025). Magyarország kibervédelmi stratégiáinak összehasonlítása és fejlődési íve 2013 és 2025 között. *Szakmai Szemle, 23*(1), 127–144. https://ojs.knbsz.gov.hu/index.php/szakmai/article/view/228

norms and governance expectations continue to be shaped by broader geopolitical competition.[66] This provides an important contextual limitation to the convergence argument developed in this dissertation: convergence is more visible within European regulatory and methodological frameworks than at the global level.

Evaluation maturity is similarly heterogeneous. Some exercises focus primarily on participant learning and qualitative feedback, while more mature programmes use predefined objectives, structured observation, After-Action Reports, improvement plans, measurable indicators, and subsequent re-testing. Documentation depth also varies considerably, particularly where exercise results are operationally sensitive.

This fragmentation means that the existence of an exercise alone says relatively little about preparedness. Two organizations may both report that they conducted a cyber exercise while testing entirely different capabilities, applying different evaluation criteria, and producing fundamentally different evidence.

However, the comparative analysis performed in the previous sections also shows that the field is not developing randomly. International methodologies such as HSEEP, ENISA exercise guidance, and BSI approaches repeatedly emphasise common principles including structured planning, objective definition, lifecycle orientation, role clarity, scenario control, systematic evaluation, After-Action reporting, corrective action, and continuous improvement.

European regulatory and policy developments reinforce several of these principles from another direction. NIS2, CER, DORA, the GDPR, the Cybersecurity Act, the EU cyber crisis-management framework, and related policy initiatives do not create one universal cyber exercise model. Nevertheless, they increasingly require organizations and authorities to demonstrate effective risk management, preparedness, continuity, resilience, incident-response capability, governance accountability, dependency management, and, in certain contexts, structured testing.

The result is an identifiable **convergence pressure**. Exercise practice remains heterogeneous, but the conditions under which exercises are considered credible and useful are becoming increasingly similar.

Dimensions of the European Convergence Framework

Based on the analysis conducted in Part I, this dissertation identifies **eight dimensions of convergence** that together form the European Convergence Framework:

1. terminology and conceptual clarity;
2. governance and accountability;
3. lifecycle orientation;
4. structured artefacts and documentation;
5. evaluation and measurement;
6. auditability and evidence;
7. dependency and ecosystem management; and
8. continuous improvement.

These dimensions do not prescribe a particular exercise scenario or technical environment. Instead, they define the common characteristics that increasingly enable exercises to become comparable, governable, and useful as resilience instruments.

[66] Molnár, D. (2022). “Great Power Cyber Diplomacy – China on the International Cyber Platform Accompanied by the United States and Russia.” Hadtudomány, 2022 special issue, 227–241.

Convergence dimension	Fragmented practice	Emerging convergence expectation	Relevance for cyber resilience
1. Terminology and conceptual clarity	Exercise, test, simulation, cyber range, stress test and related concepts are often used inconsistently.	Exercise type, purpose, scope and assessed capability are explicitly defined.	Enables meaningful comparison and prevents fundamentally different activities from being treated as equivalent.
2. Governance and accountability	Responsibilities may depend on local practice and individual planners.	Organisers, planners, players, controllers, evaluators and decision authorities have defined responsibilities.	Makes exercise preparation, execution and findings accountable and repeatable.
3. Lifecycle orientation	Exercises may be organised as isolated events.	Exercises follow a structured lifecycle from initiation and design through execution, evaluation and follow-up.	Connects preparation, execution, evaluation and improvement within one capability-development process.
4. Structured artefacts and documentation	Exercise documents differ substantially in content and quality.	Core artefacts such as objectives, scope, scenario structures, inject/MSEL logic, evaluation plans and AARs are documented systematically.	Supports repeatability, institutional memory and comparison across exercise cycles.
5. Evaluation and measurement	Success may be based mainly on participant perception or informal lessons learned.	Objectives and expected capabilities are connected to observable evidence, evaluation criteria and, where appropriate, indicators or KPIs.	Enables exercise outcomes to support evidence-based resilience assessment.
6. Auditability and evidence	Decisions, assumptions and findings may not be traceable.	Exercise design decisions, observations, findings and improvement measures are supported by traceable evidence.	Allows exercise outputs to support governance, assurance and regulatory oversight.
7. Dependency and ecosystem management	Exercises may focus primarily on one organization and its internal systems.	Scenarios increasingly address suppliers, ICT third parties, public-private coordination, critical dependencies and cascading effects.	Reflects the systemic nature of contemporary cyber crises and national resilience.
8. Continuous improvement	Exercises may end with lessons identified but no systematic follow-up.	Findings are converted into corrective actions, ownership, deadlines and subsequent verification or re-testing.	Transforms exercises from individual events into a cumulative capability-development cycle.

TABLE 4 - EUROPEAN CONVERGENCE FRAMEWORK FOR CXS

The eight dimensions provide a structured mechanism for comparing heterogeneous exercise approaches without assuming that their operational content must be identical. An executive tabletop exercise and a live-fire cyber defence exercise may differ completely in execution while still being assessed against common questions: Were the objectives defined? Were responsibilities clear? Was evidence collected? Were findings documented? Were actions assigned? Will remediation be verified?

This distinction is central to the framework. **Comparability does not require identical exercises; it requires comparable metadata, governance principles, artefacts, and evaluation logic.**

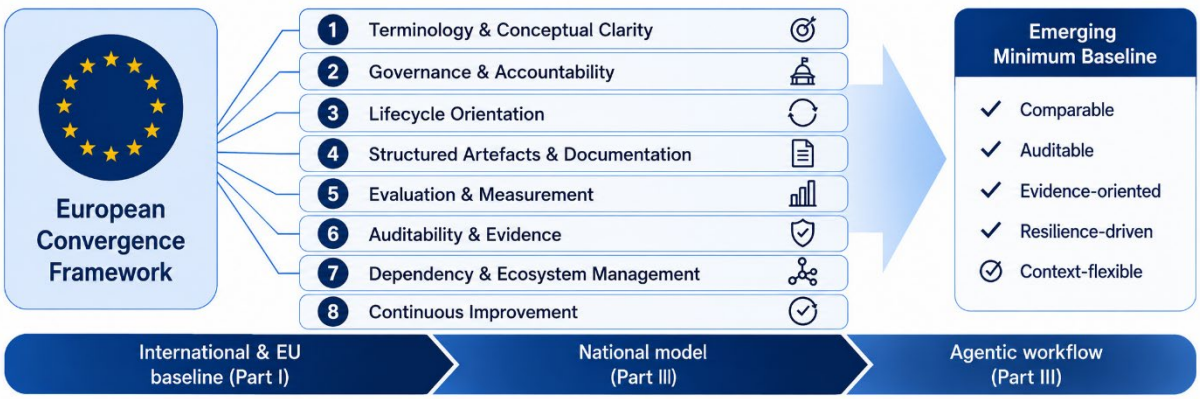


FIGURE 15 - EUROPEAN CONVERGENCE FRAMEWORK FOR CYBER EXERCISES (C1)

Convergence in Exercise Purpose: From Training to Resilience Validation

One of the strongest trends identified in this research concerns the changing purpose of cyber exercises. Historically, exercises have often been framed primarily as training and awareness activities. Training remains an important purpose, but the European resilience environment increasingly adds a second function: **validation**.

Under this logic, an exercise should not only help participants practise what they would do during a cyber incident. It should also provide evidence about whether existing capabilities, procedures, responsibilities, communication channels, and recovery arrangements function under realistic pressure.

This distinction is significant.

A training-oriented question asks: *Did the participants learn from the exercise?*

A resilience-oriented question additionally asks: *Did the exercise provide evidence that the organization or ecosystem can perform the required capability?*

The latter requires greater planning discipline. Objectives must be defined before the exercise, because evidence cannot be interpreted meaningfully without knowing what was intended to be tested. Evaluation therefore needs to be designed alongside the scenario rather than after execution. Relevant observations need to be captured systematically, and findings need to be connected to actions.

The European Convergence Framework consequently positions cyber exercises on a continuum:

training → capability validation → evidence generation → improvement → re-testing.

This does not transform every exercise into a formal audit. Rather, it demonstrates that the value of exercising increasingly depends on the ability to translate participant activity into structured evidence and subsequent learning.

Convergence in Lifecycle and Artefact Logic

A further convergence dimension concerns the increasing use of lifecycle-based exercise management.

The methodologies analysed in this dissertation differ in terminology and structure, but they share a common underlying logic. Exercises begin by determining purpose and objectives; they proceed through design and preparation; the scenario is executed under controlled conditions; observations are evaluated; and the resulting findings should lead to improvement.

This can be represented at a high level as:

Purpose → Objectives → Design → Preparation → Execution → Evidence → Evaluation → Improvement → Re-testing

The increasing emphasis on lifecycle orientation is important because it changes the exercise from an event into a **managed process**.

This development also produces convergence around exercise artefacts. Although exact templates differ, mature methodologies commonly require structured outputs such as:

- purpose and scope;
- defined exercise objectives;
- participant and governance roles;
- scenario architecture;
- inject or MSEL structures;
- communication arrangements;
- evaluation criteria and observation plans;
- evidence records;
- After-Action Reports;
- improvement actions; and
- follow-up or verification activities.

The European Convergence Framework does not require every exercise to produce identical documentation. A small tabletop exercise should not be subject to the same documentation burden as a multi-sector national cyber crisis exercise. The principle is instead **proportional standardisation**: the complexity of the artefacts should reflect the scale, purpose, risk and assurance requirements of the exercise, while a common minimum information structure remains available for comparison and programme governance.

This finding becomes particularly important for Part II of the dissertation. If exercises are to contribute to national capability steering, national actors require a minimum set of comparable information even where the underlying exercise artefacts remain confidential.

Regulatory Convergence as a Driver of Exercise Design

European convergence is also driven by the changing regulatory environment.

The regulatory instruments analysed in Section 2.4 differ in legal nature, sectoral scope, competent authorities, and specific obligations. They should therefore not be interpreted as creating one unified legal requirement to conduct a particular type of cyber exercise.

Their relevance lies instead in the **shared capability expectations** they create.

Across the regulatory landscape, recurring themes include:

- cybersecurity risk management;
- management accountability;
- incident preparedness and response;
- business continuity and recovery;
- operational resilience;

- dependency and third-party management;
- crisis coordination;
- documentation;
- evidence of implementation;
- testing and validation; and
- continuous improvement.

These themes increasingly influence exercise design even where the underlying legal instrument does not explicitly prescribe a particular exercise format. For example, if an organization is expected to maintain effective continuity arrangements, an exercise provides a mechanism to verify whether those arrangements remain usable under disruption. If third-party ICT risk is significant, meaningful exercise design increasingly requires the inclusion of relevant external dependencies. If rapid incident escalation and reporting are expected, an exercise can test whether internal detection and decision processes allow those obligations to be fulfilled under time pressure.

Consequently, regulation affects exercises in two ways. The first is **direct regulatory influence**, where testing or exercising is explicitly required or strongly embedded in the relevant resilience framework. The second is **indirect regulatory influence**, where an organization must demonstrate a capability - such as recovery, governance or incident management - and exercising becomes an appropriate mechanism for validating it. The convergence identified in this dissertation therefore concerns not only exercise methodology but also the increasing interaction between **exercise design and resilience governance**.

The Emerging Minimum Baseline

The European Convergence Framework makes it possible to derive an **emerging minimum baseline** for credible cyber exercise practice.

The term *minimum baseline* is deliberately used instead of *uniform European standard*. The analysis does not demonstrate that Europe has achieved full harmonisation, nor does it suggest that all Member States or sectors currently operate at the same maturity level. Rather, it identifies a set of characteristics that increasingly appear necessary if an exercise is expected to support more than local training.

A credible exercise intended to contribute to organizational, sectoral or national resilience should therefore increasingly be able to demonstrate at least the following:

1. Defined purpose.
The reason for conducting the exercise and the capability to be tested should be explicit.

2. Defined exercise type and scope.
The exercise format, participating community, systems, processes, organizational boundaries, assumptions and exclusions should be documented sufficiently to allow interpretation of the results.

3. Defined responsibilities.
Planning, control, participation, evaluation and decision-making responsibilities should be clear.

4. Objective-driven scenario design.
Scenario events and injects should serve defined exercise objectives rather than exist only for realism or narrative complexity.

5. Planned evaluation.
The evidence required to assess objectives should be considered during design and preparation, not only after execution.

6. Structured documentation and evidence.
Key design decisions, exercise events, observations and findings should be documented at a level appropriate to the sensitivity and scale of the exercise.

7. Relevant dependency coverage.
Where critical services depend on third parties, suppliers or other organizations, exercise design should consider these dependencies when they are relevant to the tested capability.

8. After-action improvement.
Findings should lead to defined corrective actions rather than remain at the level of lessons identified.

9. Ownership and accountability.
Improvement actions should have responsible owners and, where appropriate, timelines or prioritisation.

10. Verification and re-testing.
Material weaknesses should be followed up through subsequent validation so that resilience improvement can be demonstrated across exercise cycles.

These elements form the practical minimum baseline derived from the convergence analysis. They are intentionally independent of specific exercise technology. A tabletop exercise, distributed functional exercise, cyber range exercise or integrated live-fire simulation may all satisfy the baseline if their objectives, governance, evidence and improvement mechanisms are appropriate for their purpose.

Convergence Does Not Mean Uniformity

An important limitation of the framework must be made explicit.

The European Convergence Framework should not be interpreted as an argument for identical cyber exercises across Europe.

Uniformity could in fact reduce exercise quality where national, sectoral or organizational risks require different approaches. Energy infrastructure, public administration, healthcare, finance and military environments operate under different legal, technical and operational constraints. Their scenarios, participants, exercise environments and evaluation priorities should therefore remain context-sensitive.

The framework instead proposes **standardisation of the structure around the exercise rather than standardisation of the exercise scenario itself**.

This means that Europe can move toward greater comparability while retaining diversity in exercise content.

What should increasingly become comparable is:

- what capability was tested;
- under which conditions;
- by which type of exercise;
- involving which actors;
- according to which objectives;
- using which evaluation logic;
- producing which categories of findings; and
- resulting in which improvement actions.

This approach also addresses confidentiality. National comparability does not require unrestricted sharing of detailed vulnerabilities, technical evidence, or sensitive After-Action Reports. Exercises can be described and aggregated through higher-level metadata and capability categories while detailed operational findings remain protected.

The European Convergence Framework therefore supports the concept of **minimum comparable structure with maximum context-sensitive implementation**.

Comparative Evidence Matrix for the European Cyber Exercise Framework (C1)

The European Cyber Exercise Convergence Framework (C1) developed in this dissertation is based on the observation that cyber exercise practices remain heterogeneous across countries, sectors, and organisations, while several international methodologies and European regulatory instruments address similar underlying requirements. These include governance responsibility, structured planning, documentation, evaluation, evidence generation, dependency management, and continuous improvement.

To substantiate these observations, a comparative document analysis was conducted using eight analytical dimensions derived from the European Cyber Exercise Convergence Framework. The analysis examines whether these dimensions are explicitly addressed, indirectly supported, or insufficiently represented within selected international exercise methodologies and European regulatory instruments.

The comparison includes three methodological references, namely the Homeland Security Exercise and Evaluation Program (HSEEP), the ENISA Cybersecurity Exercise Methodology (2026), and the German Federal Office for Information Security's (BSI) business continuity and exercise guidance. These are complemented by six European regulatory instruments: the NIS2 Directive, the Critical Entities Resilience Directive (CER), the Digital Operational Resilience Act (DORA), the General Data Protection Regulation (GDPR), the EU Artificial Intelligence Act, and the EU Cyber Solidarity Act.

The sources were selected to represent different but complementary regulatory and methodological perspectives. HSEEP, ENISA, and BSI provide process-oriented exercise or continuity-testing methodologies, whereas European regulatory instruments establish requirements related to organisational preparedness, cybersecurity governance, resilience testing, accountability, and cooperation.

The objective of the analysis is not to demonstrate that these instruments are interchangeable or that they prescribe identical exercise procedures. Instead, the comparison aims to identify recurring principles that may support the development of a common European baseline for planning, documenting, evaluating, and improving cyber exercises.

Coding methodology

The assessment uses three qualitative evidence categories.

Code	Classification	Operational definition
E	Explicit evidence	The source directly addresses the relevant dimension through an identifiable methodological provision, regulatory obligation, or defined procedure.
P	Partial or indirect evidence	The source addresses an associated principle but does not fully specify its application to cyber exercises.
N	No sufficient evidence	The reviewed provisions do not establish a sufficiently clear relationship with the dimension.

The classification reflects the strength of the identified documentary relationship, not the maturity or effectiveness of an organisation, country, or regulatory framework. In particular, an explicit provision in GDPR or the AI Act does not automatically establish a requirement for a cyber exercise. The specific legal scope must be considered independently.

Comparative Evidence Matrix

Table 5 and Table 6. Evidence mapping of the eight convergence dimensions across selected methodologies and European regulatory instruments.

E = explicit evidence; P = partial/indirect evidence; N = no sufficient evidence. Switch between the methodological and regulatory sources to view all nine columns.

Convergence dimension	**HSEEP**	**ENISA**	**BSI**
1. Terminology and conceptual clarity	E	E	E
2. Governance and accountability	E	E	E
3. Lifecycle orientation	E	E	E
4. Structured artefacts and documentation	E	E	E
5. Evaluation and measurement	E	E	E
6. Auditability and evidence	E	E	E
7. Dependency and ecosystem management	P	E	P
8. Continuous improvement	E	E	E

TABLE 5 - COMPARATIVE EVIDENCE MATRIX: EXERCISE METHODOLOGIES

Convergence dimension	**NIS2**	**CER**	**DORA**	**GDPR**	**AI Act**	**Cyber Solidarity Act**
1. Terminology and conceptual clarity	P	P	E	N	N	P
2. Governance and accountability	E	E	E	E	E	E
3. Lifecycle orientation	P	P	P	N	N	P
4. Structured artefacts and documentation	P	P	E	P	P	P
5. Evaluation and measurement	E	P	E	P	P	P
6. Auditability and evidence	P	P	E	E	E	P

7. Dependency and ecosystem management	E	E	E	P	P	E
8. Continuous improvement	P	P	E	P	P	P

TABLE 6 - COMPARATIVE EVIDENCE MATRIX: EU LEGISLATION

Interpretation of the comparative matrix

The mapping identifies particularly strong methodological correspondence across HSEEP, ENISA, and BSI. All three sources address exercise organisation, defined responsibilities, structured processes, documentation, evaluation, and the subsequent improvement of preparedness arrangements. This provides direct support for the existence of common methodological principles across different institutional traditions.

The comparison of European regulatory instruments produces a more differentiated result. NIS2 establishes national cybersecurity crisis-management responsibilities, preparedness arrangements, and cybersecurity risk-management requirements, while CER incorporates resilience planning, training, and exercises into critical-entity obligations. DORA provides particularly concrete requirements concerning digital operational resilience testing, testing programmes, remediation, and validation.

GDPR and the AI Act contribute to the analysis primarily through security, governance, accountability, documentation, and verification requirements. Their relevance to cyber exercise design is therefore more indirect and depends on the systems, personal data, and AI applications involved. The Cyber Solidarity Act contributes evidence concerning coordinated preparedness testing, exercises, cross-border cooperation, and collective response capacity, without establishing a single standardised cyber exercise lifecycle.

The evidence consequently supports convergence at the level of selected principles and governance expectations, while demonstrating that methodological completeness, legal specificity, and exercise-related obligations differ considerably between sources.

1.5.1.1 *Provision-level Evidence Register*

The following register documents the rationale for the assessments presented in Table X. References to D1–D8 correspond to the eight convergence dimensions in the same order as the matrix.

HSEEP – Homeland Security Exercise and Evaluation Program (2020)

Primary source: Federal Emergency Management Agency, Homeland Security Exercise and Evaluation Program.

Dimension	Code	Evidence and justification
D1 – Terminology	E	HSEEP distinguishes exercise types, programme activities, design activities, evaluation and improvement planning through established terminology.
D2 – Governance	E	Exercise programme management identifies planning responsibilities, stakeholders, programme priorities and exercise-management functions.
D3 – Lifecycle	E	HSEEP establishes programme management, design and development, conduct, evaluation, and improvement planning as connected processes.

D4 – Artefacts	E	Standard templates include exercise plans, facilitator guides, evaluation plans, Master Scenario Event Lists and after-action documents.
D5 – Evaluation	E	Exercise Evaluation Guides link objectives and capabilities to observable tasks, performance targets and evaluation criteria.
D6 – Auditability	E	Structured observation, evaluation documentation, and After-Action Reports/Improvement Plans establish an evidence trail for exercise findings.
D7 – Dependencies	P	HSEEP supports whole-community participation and inter-organisational coordination, but dependency modelling is not universally prescribed as a core requirement for every exercise.
D8 – Improvement	E	Corrective-action planning and monitoring are integral to the exercise programme and support subsequent preparedness development.

Assessment: HSEEP provides strong evidence for seven of the eight dimensions. Its most important contribution to C1 is the demonstration that exercise planning, evaluation and improvement can be organised as a consistent programme-level methodology. Its inclusion represents an international methodological reference rather than evidence of an EU regulatory requirement.

ENISA Cybersecurity Exercise Methodology (2026)

Primary source: ENISA, The ENISA Cybersecurity Exercise Methodology, Version 1.0, February 2026.

Dimension	Code	Evidence and justification
D1 – Terminology	E	The methodology defines exercise terminology, roles, types and deliverables, supplemented by a glossary in Annex A.
D2 – Governance	E	Initiation, Design and Preparation define stakeholder participation and planning responsibilities; phase transitions include go/no-go checkpoints.
D3 – Lifecycle	E	Six explicit phases: Initiation, Design, Preparation, Execution, Evaluation and Moving Forward.
D4 – Artefacts	E	Phase-specific deliverables and templates support exercise planning, scenario design, evaluation and reporting.
D5 – Evaluation	E	Section 3.2 defines the evaluation strategy, while Chapter 5 addresses analysis, assessment and reporting.
D6 – Auditability	E	Evidence collection, documented observations and structured reporting are integrated into evaluation planning and execution.
D7 – Dependencies	E	Section 2.2 addresses system interdependencies, relevant stakeholders, supplier involvement and exercise scope boundaries.
D8 – Improvement	E	Chapter 6 establishes dissemination of results, corrective-action planning and monitoring of improvement progress.

Assessment: ENISA provides explicit methodological support for all eight convergence dimensions. It is particularly important for C1 because it consolidates exercise quality principles into a coherent European methodology, while allowing organisations and national authorities to adapt exercise formats to their own context.

ENISA's methodological guidance should not, however, be characterised as directly binding legislation. Its contribution is methodological standardisation rather than the imposition of one legally mandatory European exercise format.

BSI Standards and Exercise Guidance

Primary sources: BSI Standard 200-4, Business Continuity Management, and BSI guidance concerning tests and exercises. BSI Standard 100-4 provides an additional historical methodological reference.

Dimension	Code	Evidence and justification
D1 – Terminology	E	BSI differentiates exercise and test activities, including the distinction between general exercise objectives and objectively measurable test outcomes.
D2 – Governance	E	Exercise preparation, execution and evaluation are assigned to responsible personnel, including organisers, exercise leaders and observers.
D3 – Lifecycle	E	BSI defines activities for planning, preparing, conducting and evaluating tests and exercises within business continuity management.
D4 – Artefacts	E	Exercise concepts, scripts, documented plans, observation records and reports form part of the exercise process.
D5 – Evaluation	E	Tests and exercises evaluate the practical suitability of continuity arrangements and emergency procedures.
D6 – Auditability	E	Observations and deficiencies are documented, evaluated and reported to responsible decision-makers.
D7 – Dependencies	P	Business continuity management addresses critical resources and external dependencies, but not every exercise is required to test the complete inter-organisational ecosystem.
D8 – Improvement	E	Findings are translated into corrective measures, and the effectiveness of changes is reviewed in subsequent tests or exercises.

Assessment: BSI provides strong evidence for the relationship between cyber exercises, business continuity and continuous improvement. Its relevance lies in the operationalisation of exercising within an established management framework. Unlike ENISA, however, BSI 200-4 is primarily concerned with business continuity management and should not be presented as an exclusively cyber-exercise-specific methodology.

NIS2 Directive – Directive (EU) 2022/2555

Relevant provisions: Articles 7, 9, 20 and 21, particularly Article 9(4) concerning national cyber crisis response plans and Article 21(2) concerning cybersecurity risk-management measures.

Dimension	Code	Evidence and justification
D1 – Terminology	P	NIS2 establishes common cybersecurity, incident-management and crisis-management terminology, but does not define a comprehensive cyber exercise taxonomy.
D2 – Governance	E	Articles 9 and 20 establish institutional and management responsibilities. Article 9(4) addresses national crisis plans and preparedness measures, including exercises and training.
D3 – Lifecycle	P	Preparedness, incident handling, continuity and crisis-management responsibilities are addressed, but no uniform end-to-end cyber exercise lifecycle is prescribed.
D4 – Artefacts	P	Risk-management policies, procedures, crisis plans and incident documentation are required, but not a common set of exercise-specific planning artefacts.
D5 – Evaluation	E	Article 21(2)(f) requires policies and procedures to assess the effectiveness of cybersecurity risk-management measures.
D6 – Auditability	P	Documented cybersecurity governance and incident-reporting requirements support accountability, but no universal exercise-level provenance or audit-log structure is prescribed.
D7 – Dependencies	E	Article 21(2)(d) explicitly addresses supply-chain security and relationships with direct suppliers and service providers.
D8 – Improvement	P	Effectiveness assessment and risk-management obligations support corrective improvement, but NIS2 does not establish a common exercise findings-to-retesting lifecycle for all entities.

Assessment: NIS2 provides substantial governance and preparedness requirements that can inform cyber exercise design. Its relevance to C1 is strongest in governance, effectiveness assessment, national coordination and supply-chain dependencies. Importantly, Article 9(4) expressly includes exercises in national preparedness arrangements. This is a stronger relationship than merely inferring exercises from general risk-management duties.

Critical Entities Resilience Directive – Directive (EU) 2022/2557

Relevant provisions: Articles 11–13 and 20, covering critical-entity risk assessment, resilience measures, documentation, training and European cooperation.

Dimension	Code	Evidence and justification
D1 – Terminology	P	CER defines concepts concerning critical entities, resilience and essential services, but does not establish a comprehensive cyber exercise terminology system.
D2 – Governance	E	The directive defines responsibilities for Member States, competent authorities and critical entities in establishing and maintaining resilience measures.
D3 – Lifecycle	P	Prevention, response, mitigation and recovery are addressed as resilience functions, but no dedicated cyber exercise lifecycle is prescribed.
D4 – Artefacts	P	Article 13(2) provides for a resilience plan or equivalent documentation, without specifying the full set of exercise-planning and evaluation artefacts.

D5 – Evaluation	P	Risk assessment, resilience measures, training and exercises support preparedness assessment, but common cyber exercise evaluation metrics are not defined.
D6 – Auditability	P	Resilience documentation and supervisory mechanisms support demonstrability, but exercise-specific provenance and evidence structures remain largely unspecified.
D7 – Dependencies	E	CER explicitly addresses essential-service dependencies, external service providers, continuity arrangements and alternative supply chains.
D8 – Improvement	P	Regular risk assessment, resilience measures and exercising support adaptation, but a uniform after-action corrective-action and re-testing process is not defined.

Assessment: CER provides strong evidence for the importance of dependency-aware exercising and resilience governance. Its contribution is particularly relevant for national and cross-sector exercises because the performance of critical entities cannot be evaluated solely through technical cybersecurity controls.

The Commission's 2026 guidance additionally discusses tabletop and operational exercises for validating resilience arrangements, reinforcing the practical relationship between critical-entity governance and exercising. This guidance is useful interpretive material, but should be distinguished from the binding directive itself.

DORA – Regulation (EU) 2022/2554

Relevant provisions: Articles 5–6, 24–27 and 28, covering ICT governance, digital operational resilience testing, threat-led penetration testing and ICT third-party risk.

Dimension	Code	Evidence and justification
D1 – Terminology	E	DORA expressly defines ICT risk-management and resilience-testing concepts, including advanced threat-led penetration testing.
D2 – Governance	E	Management-body responsibilities, testing governance, independent testing arrangements and defined testing responsibilities are established.
D3 – Lifecycle	P	DORA establishes recurring testing programmes and structured advanced testing activities, but not one general cyber exercise lifecycle covering every exercise type.
D4 – Artefacts	E	Testing programmes, documented test results, findings and advanced TLPT documentation provide explicit testing-related artefacts.
D5 – Evaluation	E	Chapter IV expressly establishes digital operational resilience testing and effectiveness assessment.
D6 – Auditability	E	Documented results, remediation arrangements, independent testing and applicable supervisory processes support traceable testing outcomes.
D7 – Dependencies	E	DORA addresses ICT third-party risk, critical or important functions and dependencies on external technology service providers.
D8 – Improvement	E	Article 24(5) expressly addresses prioritisation, classification, remediation and internal validation of identified weaknesses.

Assessment: DORA provides the strongest sector-specific regulatory example of structured resilience testing among the selected instruments. It demonstrates that testing can be integrated into formal governance, documentation, remediation and assurance requirements. However, DORA concerns the financial sector, and its testing provisions do not establish a universal cyber exercise obligation for every organisation.

GDPR – Regulation (EU) 2016/679

Relevant provisions: Articles 5(2), 24, 28, 30 and 32, addressing accountability, processor relationships, processing records, information security and testing effectiveness.

Dimension	Code	Evidence and justification
D1 – Terminology	N	GDPR does not define a cyber exercise taxonomy or exercise-specific terminology.
D2 – Governance	E	GDPR establishes explicit accountability and responsibilities for controllers and processors.
D3 – Lifecycle	N	No cyber exercise lifecycle is established.
D4 – Artefacts	P	Documentation and processing-record requirements may affect exercise data handling but are not exercise-planning artefact requirements.
D5 – Evaluation	P	Article 32(1)(d) requires regular testing, assessment and evaluation of security measures; this does not specifically require cyber exercises.
D6 – Auditability	E	Article 5(2) expressly establishes accountability, supported by demonstrability and record-related provisions elsewhere in the regulation.
D7 – Dependencies	P	Controller–processor relationships are relevant to third-party data handling but do not amount to a comprehensive cyber resilience dependency model.
D8 – Improvement	P	Regular assessment of security effectiveness can inform corrective improvement, but an exercise-specific after-action and re-testing process is not established.

Assessment: GDPR contributes primarily to the governance and accountability dimensions. Its relevance for cyber exercises concerns lawful processing of exercise data, confidentiality, documentation and the evaluation of relevant security measures. It should not be interpreted as prescribing a mandatory cybersecurity exercise methodology.

EU Artificial Intelligence Act – Regulation (EU) 2024/1689

Relevant provisions: Articles 9–15, 16, 26 and 72, insofar as applicable, particularly requirements concerning high-risk AI systems, technical documentation, logging, human oversight and monitoring.

Dimension	Code	Evidence and justification
D1 – Terminology	N	The AI Act defines AI-related concepts but does not establish a cyber exercise classification system.
D2 – Governance	E	Applicable requirements define responsibilities concerning risk management, providers, deployers and human oversight.

D3 – Lifecycle	N	AI-system lifecycle governance does not constitute a cyber exercise lifecycle.
D4 – Artefacts	P	Technical documentation and instructions for applicable AI systems can inform AI-supported exercise artefact governance, but are not exercise-documentation requirements.
D5 – Evaluation	P	Testing, validation, accuracy, robustness and monitoring requirements concern regulated AI systems rather than the evaluation of cyber exercises.
D6 – Auditability	E	Article 12 establishes logging requirements for high-risk AI systems, supported by technical documentation and relevant monitoring obligations.
D7 – Dependencies	P	Provider–deployer relationships and AI value-chain responsibilities may inform governance of AI-supported exercise tools, but do not establish a cyber exercise dependency model.
D8 – Improvement	P	Risk management and monitoring support corrective intervention for regulated AI systems, but not a general cyber exercise improvement cycle.

Assessment: The AI Act supports the wider argument that responsible AI integration requires governance, traceability and human oversight. Nevertheless, its provisions must be interpreted according to their legal scope. An AI-supported exercise-planning tool is not automatically a high-risk AI system merely because it is used in a cybersecurity or defence-related context.

The relevance of these provisions for C1 should therefore be presented as a conditional governance connection, rather than a direct legal obligation to operate cyber exercises through an auditable agentic workflow.

EU Cyber Solidarity Act – Regulation (EU) 2025/38

Relevant provisions: Articles 11–14 and the provisions concerning coordinated preparedness, cross-border cooperation and incident review mechanisms.

Dimension	Code	Evidence and justification
D1 – Terminology	P	The Act establishes terminology concerning preparedness testing, cybersecurity emergencies and coordinated response, without defining a common exercise taxonomy.
D2 – Governance	E	EU-level and Member State coordination arrangements, cybersecurity emergency mechanisms and reserve services establish explicit institutional responsibilities.
D3 – Lifecycle	P	Preparedness, response and incident review functions address related stages, but no complete exercise lifecycle is specified.
D4 – Artefacts	P	Coordinated testing and associated administrative arrangements require structured information, without prescribing a comprehensive standard exercise artefact set.
D5 – Evaluation	P	Coordinated preparedness testing supports capability assessment, but does not establish a single exercise evaluation methodology or common KPI set.
D6 – Auditability	P	Testing arrangements and incident review support accountability, but uniform exercise-level evidence and provenance requirements are not prescribed.

D7 – Dependencies	E	Cross-border collaboration, shared capabilities and mutual assistance explicitly address resilience beyond individual organisations or Member States.
D8 – Improvement	P	Preparedness support and incident review can inform future measures, but the Act does not prescribe a uniform exercise corrective-action and re-testing cycle.

Assessment: The Cyber Solidarity Act provides important evidence of European-level preparedness cooperation and testing support. Articles 12 and 13 distinguish voluntary coordinated testing from additional supported preparedness activities, including exercises and training. The Act supports the argument that cyber resilience increasingly requires coordinated mechanisms beyond individual organisations, while leaving exercise design methods largely open.

Cross-Dimensional Analysis of the Findings

The comparative analysis provides evidence of recurring methodological and regulatory principles, while also demonstrating that the strength and nature of alignment differ across the eight dimensions.

Terminology and conceptual clarity

Terminological standardisation is strongest in the methodological sources. HSEEP, ENISA and BSI differentiate exercise types, planning activities, evaluation procedures and supporting documents. DORA also provides specific terminology for advanced resilience testing.

However, terminology remains fragmented across the broader regulatory environment. Instruments such as GDPR and the AI Act define concepts within their respective legal domains rather than providing a common vocabulary for cyber exercising. This indicates that a European cyber exercise taxonomy would require an explicit synthesis of existing terminologies rather than the simple adoption of one regulatory vocabulary.

Governance and accountability

Governance and accountability demonstrate the strongest cross-source correspondence. All nine instruments establish identifiable governance responsibilities, although their institutional and legal purposes differ.

Methodological sources define responsibilities for exercise organisers, planners, controllers and evaluators. Regulatory instruments establish obligations for management bodies, competent authorities, critical entities, controllers, processors, or AI-system actors.

For the European Convergence Framework, this suggests that clearly defined responsibilities and accountable decision-making constitute a particularly strong common principle, even where the actors and legal obligations differ.

Lifecycle orientation

Lifecycle orientation is clearly established in HSEEP, ENISA and BSI. Each describes exercising as a structured process encompassing preparation, execution, evaluation and subsequent improvement.

European legislation is less prescriptive concerning exercise-specific lifecycle structures. Several regulatory instruments address preparedness, testing, response, recovery and monitoring, but these should not be equated with the lifecycle of a cyber exercise.

The evidence therefore supports a methodological common baseline, but not the existence of a single legally prescribed European cyber exercise lifecycle.

Structured artefacts and documentation

Documentation is consistently addressed across the selected sources, although its purpose differs substantially.

HSEEP, ENISA and BSI identify exercise-specific documents, including exercise plans, scenarios, control materials and evaluation reports. DORA also establishes concrete testing-related documentation requirements within its regulatory scope.

Other instruments contribute more general documentation obligations, such as cybersecurity policies, resilience plans, technical documentation and processing records.

These findings support standardisation of a minimum exercise artefact set as a proposed design response, while recognising that the same artefacts are not universally required by EU law.

Evaluation and measurement

Evaluation and measurement show considerable correspondence across the methodologies and selected regulations. The strongest evidence concerns the need to assess whether plans, capabilities, controls or resilience arrangements function effectively.

The principal limitation is that evaluation does not always mean the same thing. Exercise evaluation, cybersecurity control effectiveness, regulatory compliance assessment and AI-system performance testing involve different objects of assessment and may require different indicators.

A common European framework should therefore distinguish a shared evaluation principle from the domain-specific metrics used to implement it.

Auditability and evidence

Auditability and evidence generation are strongly represented, particularly within HSEEP, ENISA, BSI, DORA, GDPR and the AI Act.

Nevertheless, the purposes of the evidence differ. Exercise methodologies focus on observations and after-action findings. DORA concerns resilience-testing outputs and supervisory assurance. GDPR establishes data-protection accountability, while the AI Act addresses documentation and logging for regulated AI systems.

These sources support the importance of traceable documentation, but they do not collectively establish one universal exercise evidence architecture.

Dependency and ecosystem management

Dependency management is particularly important in national cyber resilience. The European regulatory instruments provide substantial evidence concerning supply chains, critical services, third-party relationships and cross-border cooperation.

The ENISA methodology additionally addresses system interdependencies, external stakeholders and the importance of defining realistic exercise scope.

This supports the inclusion of dependency and ecosystem management as an independent dimension of C1. It also establishes a direct connection to the national governance model, where exercises must account for inter-organisational and cross-sector dependencies.

Continuous improvement is explicit within the three methodological references and DORA. Other regulations also establish mechanisms supporting reassessment, monitoring or adaptation, although they do not necessarily prescribe exercise-specific improvement cycles.

A significant distinction must therefore be maintained between identifying deficiencies, implementing corrective action and demonstrating verified improvement. The existence of a documented improvement requirement does not, by itself, prove that a corrective action has been effective.

The framework consequently identifies documented corrective-action tracking and subsequent verification as desirable requirements for a national cyber exercise programme, rather than claiming that all selected instruments already mandate this complete process.

Consolidated Evidence Assessment

Distribution of qualitative evidence classifications by convergence dimension.

Convergence dimension	*Explicit (E)*	*Partial (P)*	*Insufficient (N)*
Terminology and conceptual clarity	4	3	2
Governance and accountability	9	0	0
Lifecycle orientation	3	4	2
Structured artefacts and documentation	4	5	0
Evaluation and measurement	5	4	0
Auditability and evidence	6	3	0
Dependency and ecosystem management	5	4	0
Continuous improvement	4	5	0
Total	*40*	*28*	*4*

TABLE 7 - EVIDENCE CLASSIFICATIONS BY CONVERGENCE DIMENSION

Number of explicitly and partially supported relationships across nine selected sources.

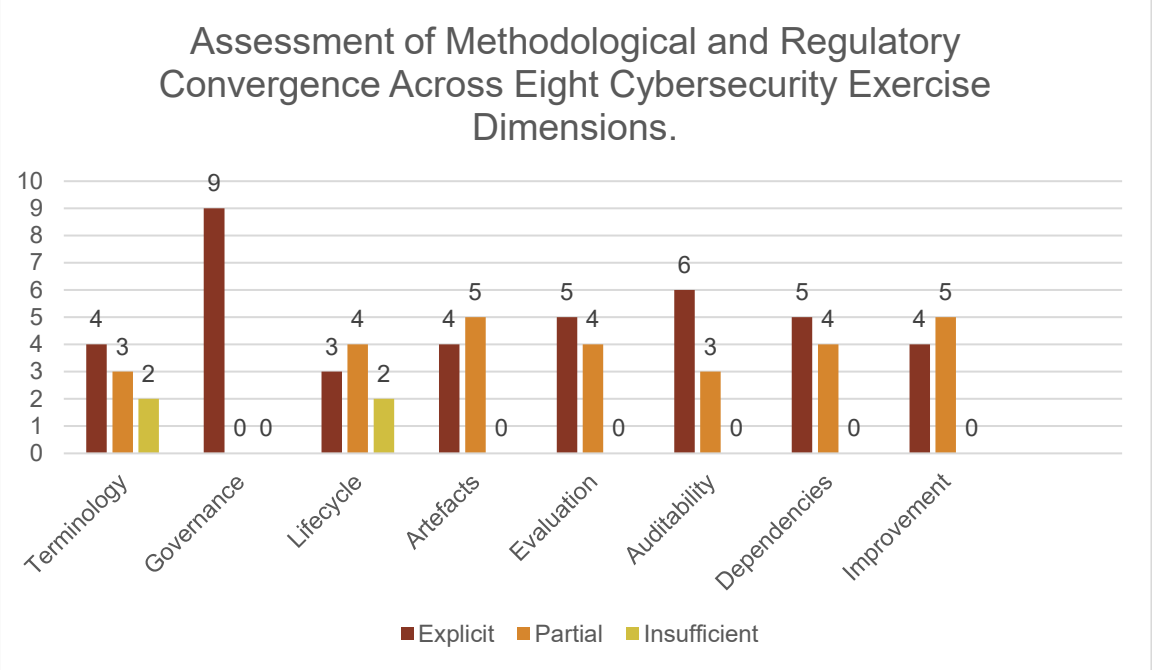


TABLE 8 - ASSESSMENT OF REGULATORY CONVERGENCE

The table presents the distribution of Explicit, Partial, and Insufficient coverage across eight cybersecurity exercise convergence dimensions: terminology, governance, lifecycle, artefacts, evaluation, auditability, dependencies, and improvement. Governance demonstrates the strongest alignment, with nine explicit assessments, while terminology and lifecycle show greater variation, indicating areas where further methodological harmonisation and standardisation may be required.

The distribution indicates that governance and accountability have the strongest documentary correspondence among the selected sources. Lifecycle orientation and terminology display more limited alignment, particularly between exercise methodologies and broader EU legislation.

These counts are descriptive coding results, not a statistically validated convergence index. All source–dimension pairs are weighted equally for descriptive purposes, despite important differences in the authority, scope and specificity of the underlying instruments. The distribution must therefore not be interpreted as measuring the degree of European harmonisation or changes in convergence over time.

Methodological Limitations and Potential Contradictions

The analysis identifies several limitations relevant to the interpretation of C1.

Differences in regulatory scope. HSEEP, ENISA and BSI provide methodological guidance for exercises or continuity testing. The selected EU regulations and directives primarily establish substantive obligations within their respective domains. Similar terminology does not imply equivalent legal requirements. In particular, resilience testing, penetration testing and cyber exercises must remain analytically distinguishable.

Differences between guidance and binding legislation. Methodological recommendations and directly applicable or nationally transposed legal obligations cannot be treated as having equivalent normative authority. The matrix identifies conceptual relationships rather than assigning identical legal force to all sources.

Different objects of evaluation. Some instruments concern organisational preparedness, others technical-system security, data protection, critical-entity resilience or AI-system performance. Their evaluation provisions may overlap conceptually while applying to different objects and requiring different evidence.

Limits of cross-sectional analysis. The comparison identifies common principles within the selected documents. It does not demonstrate that convergence has increased over a defined historical period. Such a claim would require longitudinal analysis using comparable documents or observations from multiple periods.

Applicability of AI regulation. The AI Act's documentation, logging and oversight requirements depend on the classification and use of the relevant AI system, as well as the applicable provisions and dates. Their presence in this matrix should not be interpreted as establishing an automatic high-risk classification for AI-supported cyber exercise planning.

Relationship Between the Framework and H1

The European Convergence Framework represents the principal analytical contribution associated with **Hypothesis H1** and **MRQ1**. H1 proposes that the international and European cyber exercise landscape remains fragmented while international best practices and European regulatory developments increasingly generate convergence towards more documented, comparable, auditable and resilience-oriented practice. The analysis in Part I supports this proposition in two stages.

First, fragmentation remains observable. Terminology, exercise formats, governance arrangements, sectoral maturity, technical infrastructures, documentation practices and evaluation methods continue to differ substantially.

Second, these differences coexist with increasingly consistent underlying expectations. International methodologies converge around lifecycle-based planning, clear objectives, structured artefacts, evaluation, After-Action reporting and improvement. European regulation and resilience policy reinforce expectations concerning governance accountability, continuity, dependency management, incident preparedness, testing, documentation and evidence.

H1 should therefore not be interpreted as claiming completed European harmonisation. The finding is more precise:

European cyber exercise practice remains heterogeneous at the implementation level, while convergence is increasingly visible at the level of governance principles, lifecycle logic, evidence expectations, and continuous improvement.

The European Convergence Framework captures this intermediate condition. Its scientific contribution lies in converting the broad observation of “increasing convergence” into an analytically structured model consisting of defined convergence dimensions and an emerging minimum exercise baseline.

Scientific Contribution and Originality

Evidence-Based Interpretation of the European Convergence Framework

The comparative evidence matrix provides a structured basis for evaluating the first research hypothesis and substantiating the proposed European Cyber Exercise Convergence Framework. The analysis demonstrates that international exercise methodologies and European regulatory instruments exhibit recurring principles concerning governance, documentation, evaluation, evidence generation, dependency management, and continuous improvement. These commonalities are particularly pronounced within established exercise methodologies, where HSEEP, ENISA and BSI provide relatively comprehensive approaches to structured planning, controlled execution, evaluation, and subsequent improvement.

The regulatory analysis reveals a more differentiated pattern. European instruments such as NIS2, CER, DORA and the Cyber Solidarity Act establish requirements and mechanisms that are relevant to preparedness, resilience testing, crisis coordination and organisational accountability. However, these instruments differ in their legal scope, regulatory objectives and degree of specificity. GDPR and the AI Act contribute primarily through governance, documentation and accountability requirements rather than direct cyber exercise obligations. Consequently, the identified alignment should be interpreted as the presence of shared methodological and governance principles rather than evidence of a uniform European cyber exercise regime.

The comparative findings support the construction of the European Cyber Exercise Convergence Framework as an analytical synthesis of eight dimensions. The framework provides a common reference structure through which heterogeneous cyber exercise approaches can be examined and described. Its contribution lies in identifying which principles are already explicitly represented in established methodologies and regulations, where only partial alignment exists, and where additional methodological development would be required to establish a common baseline.

The analysis also highlights an important distinction between methodological convergence and regulatory harmonisation. While several sources establish comparable expectations regarding accountability, documentation, assessment and improvement, they do not necessarily require identical exercise processes, artefacts or evaluation indicators. The framework therefore does not propose uniformity of exercise design. Instead, it establishes a structured basis for comparability

while preserving legitimate differences in exercise objectives, institutional responsibilities, sectoral requirements and operational contexts.

In relation to Hypothesis H1, the evidence supports the proposition that recurring methodological principles and regulatory requirements can be identified despite the continued fragmentation of international and European cyber exercise practices. The findings do not, however, establish that convergence has increased over time, since the research does not include a longitudinal measurement of methodological or regulatory change. The scientific contribution should therefore be understood as the systematic identification and consolidation of common principles and areas of partial alignment.

This result provides the analytical foundation for the subsequent national research layer. The European Convergence Framework identifies the common characteristics that may inform national cyber exercise governance, while the National Cyber Exercise Governance Model translates these characteristics into programme structures, institutional responsibilities, comparable artefacts, minimum datasets, measurement mechanisms and improvement processes. In this way, C1 establishes the conceptual baseline for C2 without assuming that European regulation alone produces nationally comparable or auditable cyber exercise practices.

The originality of the European Convergence Framework does not lie in proposing individual principles such as clear objectives, structured evaluation or after-action learning. These principles already exist in recognised methodologies and professional practice.

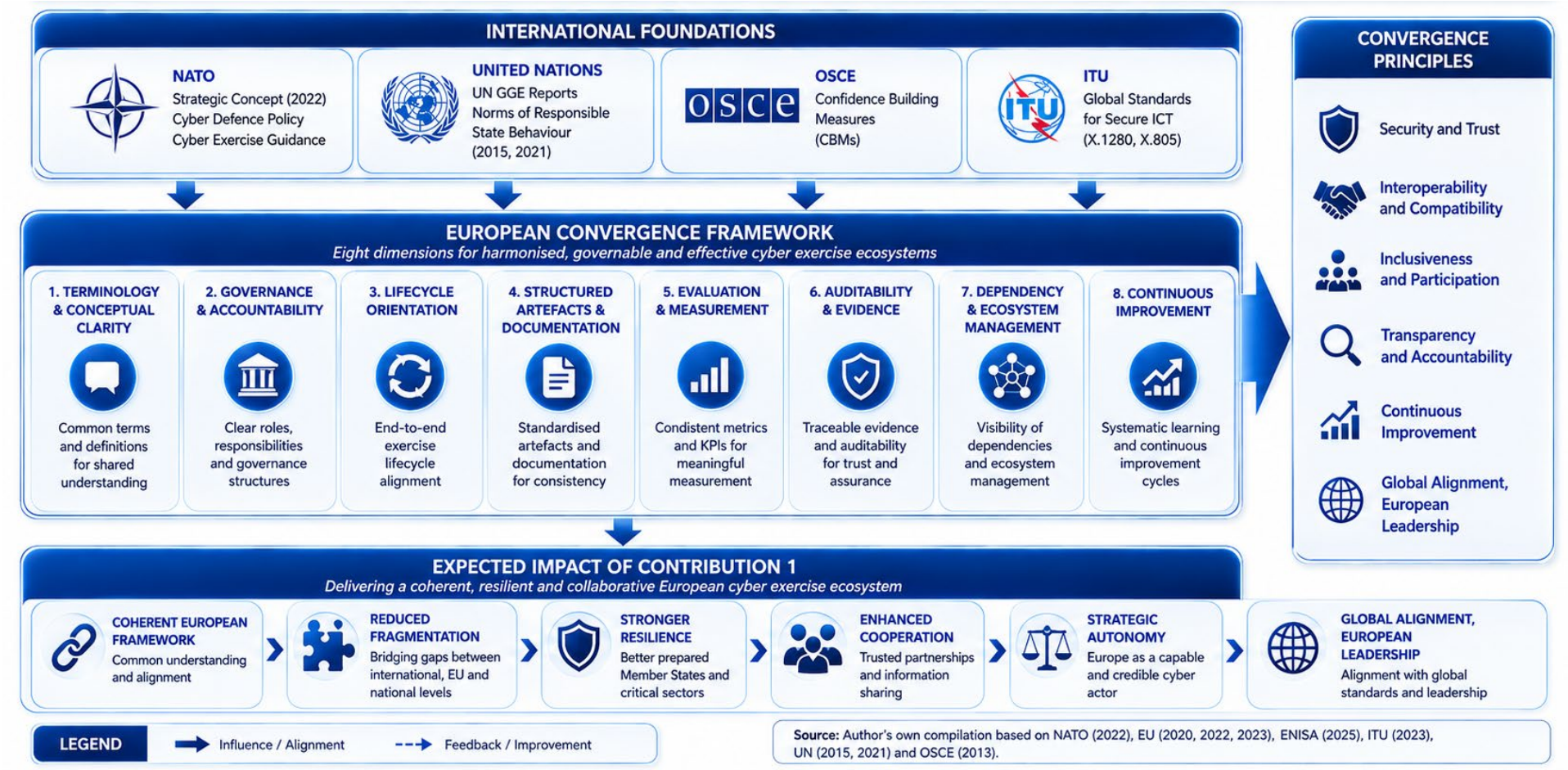


FIGURE 16 - CONTRIBUTION 1 : CONVERGENCE FRAMEWORK

The scientific contribution lies in their **cross-framework and cross-regulatory synthesis** (see overview in Figure 16.)

Existing methodologies generally explain how to conduct exercises. Regulatory instruments define resilience, governance or testing expectations within particular legal and sectoral contexts. National exercise programmes demonstrate how exercising is implemented in practice. These sources are typically considered separately.

The European Convergence Framework brings these perspectives together and asks a different question:

Which common characteristics emerge when international exercise methodology, European resilience regulation, and contemporary exercise practice are analysed as parts of the same cyber resilience ecosystem?

The resulting framework identifies the shared structural dimensions that connect these domains.This has three consequences.

> First, it provides an analytical baseline for comparing heterogeneous exercise practices without requiring identical exercise formats.
>
> Second, it creates a bridge between **regulatory expectations and exercise engineering**, making visible how high-level resilience requirements can influence objectives, scenario scope, evidence collection and improvement mechanisms.
>
> Third, it establishes the external reference point required for the constructive contribution developed in Part II. Once an emerging international and European baseline has been identified, the next research problem becomes how a nation state can translate this baseline into coherent governance, comparable exercise data, programme-level measurement and continuous improvement.

The contribution of Part I can therefore be summarised as follows:

European Convergence Framework: a structured synthesis of international methodologies, European regulatory drivers, and exercise practices that identifies eight emerging convergence dimensions and derives a minimum baseline for comparable, auditable and resilience-oriented cyber exercises without requiring uniform exercise implementation.

From European Convergence to National Governance

The European Convergence Framework also marks the transition from the analytical to the constructive part of the dissertation.

Part I establishes **what is converging**. It identifies the emerging expectations that increasingly characterise mature cyber exercise practice: conceptual clarity, governance, lifecycle orientation, structured artefacts, evaluation, auditability, dependency management and continuous improvement.

However, identifying these expectations does not explain how they should be implemented at national level.

European requirements are applied within national institutional environments containing different authorities, sectoral structures, legal responsibilities, exercise capabilities, technical infrastructures and confidentiality constraints. Consequently, convergence at European level does not automatically create coherent national cyber exercise governance.

This creates the research problem addressed by Part II.

The national challenge is to translate the convergence baseline into an operating capability system that determines:

- who governs national exercising;
- which exercises should form part of the national portfolio;
- what information should be collected consistently;
- how different exercises can be compared without oversimplification;
- how evaluation results can support national steering;
- how sensitive information can be protected;
- how improvement actions can be monitored;
- and how weaknesses can be re-tested across subsequent exercise cycles.

The European Convergence Framework therefore functions as the **external baseline** of the dissertation, while the National Cyber Exercise Governance Model developed in Part II represents its **national implementation logic**.

The transition between the two can be expressed as follows:

International methodologies and European requirements
→ European Convergence Framework
→ national translation and governance
→ comparable cyber exercise programme
→ evidence and continuous resilience improvement.

In this way, Part I answers the question of **what common expectations are emerging**, while Part II addresses **how these expectations can be transformed into a governable national capability system**.

1.6 Regulatory pathway: convergence trends and mandatory CX

There is a trend across EU regulations toward operationally validated resilience. Even though legal documents like NIS2 and CER don't require specific cyber exercises, they are putting more and more pressure on organizations to show that their plans for risk management, incident response, and business continuity not only exist but actually operate when things go wrong. This change turns cyber exercises from a voluntary way to show that you're ready to a practical way to show that you're ready and to keep getting better. DORA, NIS2, and CER are three things that are driving this trend: DORA makes operational resilience testing in the financial sector more formal; NIS2 raises the bar for risk-based security management and governance; and CER ties resilience to an all-hazards, essential-services view. They all point to a European direction of travel where exercises help people meet compliance standards and show that they can do things in the real world.

This convergence is also visible in the themes that exercises increasingly need to cover across sectors. The common focus is changing from just technical performance (such detection and containment) to outcomes that keep services running and coordination between different organizations. In practice, this means not just using the SOC and incident response protocols, but also making executive decisions, communicating during a crisis, following escalation paths, and setting recovery priorities, especially when important services or functions are at danger. Managing dependencies, especially those involving ICT third parties and supply chains, is another important issue that runs across the whole thing. DORA puts third-party risk front and center; CER focuses on cascading disruptions across interconnected infrastructures; and NIS2's focus on risk management makes these kinds of interconnections a natural choice for scenario selection. So, exercises are putting more stress on things like coordinating with suppliers, escalating contracts,

exchanging information, and working together to recover, instead than dismissing outside partners as unimportant.

Simultaneously, there is a gradual methodological convergence between civilian and defence-oriented exercise activities, notwithstanding their distinct purposes and limitations. Military cyber exercises usually focus on threats from other countries and strategic defence goals, and they generally take place in secret. Civilian exercises focus on making sure that important services are available, that they don't have a negative effect on society, and that they follow the rules. Even though these two areas are different, they both depend on the same basic ideas: standardized roles, realistic scenarios and inject control, structured evaluation, and learning loops after the action. This is in line with the larger European trend toward more systematic and comparable exercise governance.

The practical effect is that future "mandatory expectations" will probably come together to form a minimal cyber exercise baseline that is the same for all sectors but has different content for each. A baseline like this would usually have a tiered exercise portfolio (with discussion-based exercises, functional drills, and periodic integrated exercises), risk-driven scoping and goals based on organizational and national risk assessments, success criteria that focus on continuity (like recovery and service availability), and an audit trail that shows what was exercised and what was learned. Most importantly, converging frameworks inherently enable a shared improvement mechanism: exercises should feed into an AAR → corrective actions → re-testing loop, which allows for observable progress over time and reduces systemic gaps that happen repeatedly.

Finally, convergence makes a compelling case for combining national exercise programs to prevent duplication. National programs can coordinate mandatory and voluntary activities within a single portfolio because CER and NIS2 obligations can often be shown through broader resilience and crisis-management exercises. They can also prioritize additional cyber exercises where they fill in gaps that sectoral or authority-led resilience exercises don't already cover. Over time, this method helps Europe move toward greater comparability and benchmarking of cyber resilience by standardizing the underlying exercise governance, assessment rationale, and improvement monitoring. This does not mean that all scenarios or full harmonization are needed.

1.7 Summarizing H1 International Convergence Hypothesis

Part I examined the international and European cyber exercise landscape, major methodologies and standards, and the regulatory and policy environment that increasingly shapes cyber resilience and preparedness. The analysis confirmed that cyber exercise practice remains heterogeneous across terminology, exercise formats, governance structures, maturity levels, documentation practices, technical environments, and evaluation approaches. At the same time, recurring principles can be identified across international methodologies and European regulatory developments.

The main analytical result of Part I is the **European Convergence Framework for Cyber Exercises**, which consolidates these recurring principles into eight convergence dimensions: terminology and conceptual clarity; governance and accountability; lifecycle orientation; structured artefacts and documentation; evaluation and measurement; auditability and evidence; dependency and ecosystem management; and continuous improvement. These dimensions represent an emerging minimum baseline for more comparable, auditable, and resilience-oriented cyber exercise practice.

The findings therefore support **H1**. The evidence does not indicate full harmonisation or the emergence of one uniform European cyber exercise model. Instead, convergence is increasingly visible at the level of common governance expectations, lifecycle logic, evidence requirements, evaluation practices, and improvement mechanisms. Different countries, sectors, and

organizations may continue to use different exercise formats and implementation approaches while operating within an increasingly similar set of underlying principles.

European regulatory and policy developments further reinforce this trend. Instruments such as NIS2, CER, DORA, the Cybersecurity Act, GDPR, the EU AI Act, and the European cyber crisis-management framework increasingly emphasise preparedness, accountability, resilience testing, evidence, dependency management, and continuous improvement. These requirements do not prescribe one universal exercise format, but they strengthen the demand for structured and demonstrable preparedness.

Part I therefore establishes the **external baseline for the national research layer**. The central question for Part II is no longer whether convergence exists, but how this emerging international and European baseline can be translated into a coherent national cyber exercise governance system. The European Convergence Framework consequently serves as the reference point for the National Cyber Exercise Governance Model developed in Part II.

In summary, H1 is supported: cyber exercise practice remains fragmented in implementation, while an increasingly common European baseline is emerging around governance, lifecycle structure, documentation, evidence, evaluation, and continuous improvement (see in Figure 17**).**

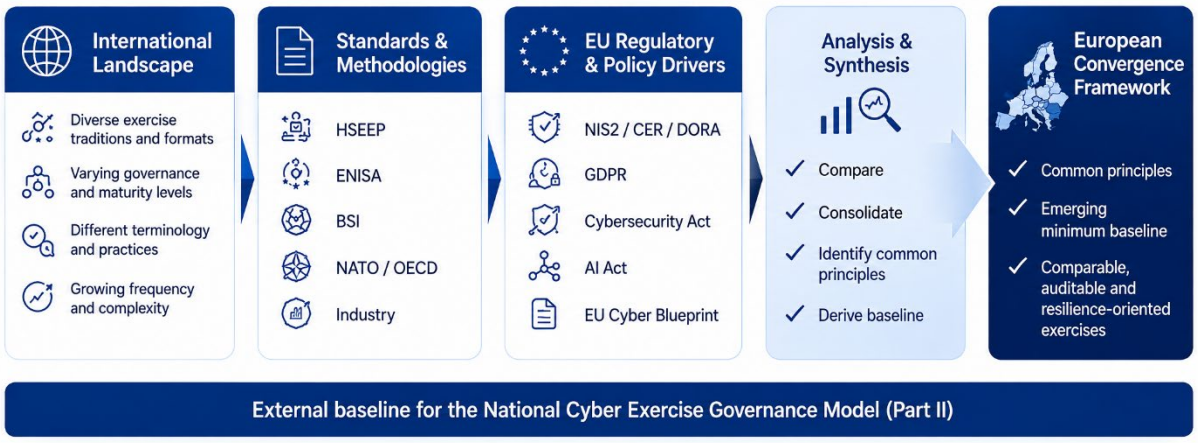


FIGURE 17 - INTERNATIONAL CONVERGENCE TO NATIONAL LAYER

2 PART II – National research layer

2.1 Cyber exercises globally, international convergence and regulatory landscape

Part II develops the **national research layer** of the dissertation and addresses the transition from the international and European convergence identified in Part I to the design of a coherent national cyber exercise governance system. Whereas Part I examined how cyber exercise practices differ internationally and which methodological, regulatory, and policy developments increasingly create a common European baseline, Part II asks how such external expectations can be translated into national structures, programme logic, common information requirements, improvement mechanisms, and repeatable exercise processes.

This transition is central to the cumulative research logic of the dissertation. The European Convergence Framework developed in Part I identifies an emerging baseline around terminology and conceptual clarity, governance and accountability, lifecycle orientation, structured artefacts, evaluation and measurement, auditability and evidence, dependency and ecosystem management, and continuous improvement. These principles, however, do not automatically become national

capability. Their practical value depends on how they are interpreted through national legislation, competent authorities, institutional responsibilities, sectoral structures, and existing exercise practices. Part II therefore examines the governance mechanisms required to convert European and international expectations into an operationally meaningful national cyber exercise ecosystem.

The national layer primarily addresses **Hypothesis H2** and **Main Research Question MRQ2**. H2 proposes that a national defence-oriented cyber exercise ecosystem based on defined governance responsibilities, common quality criteria, comparable artefacts and datasets, coordinated exercise programmes, a structured lifecycle, and mandatory improvement loops provides a more suitable framework for comparability, auditability, and resilience-oriented improvement than fragmented or ad hoc exercise activity. MRQ2 consequently asks how such a national ecosystem can be designed and governed so that international and European requirements are translated into comparable, auditable, and continuously improving national cyber exercise practice.

A central argument of Part II is that the national challenge is not simply to conduct more exercises. National resilience cannot be inferred from exercise frequency alone. Organizations may participate in numerous tabletop exercises, technical cyber-range activities, sectoral drills, crisis-management simulations, or multinational exercise series, while their results remain difficult to compare and their findings remain weakly connected to long-term improvement. The national research layer therefore treats cyber exercises as components of a broader **capability-development and governance system**, rather than as isolated events.

Recent Hungarian national-security reserach provides **an additional perspective on the institutional implications of artificial intelligence. The multidisciplinary research collected by Kenedli and Magyar (2025) situates AI within a broader national-security context that extends beyond purely technical implementation**.[67] Building on this perspective, the present dissertation examines how emerging technological capabilities can be integrated into national cyber exercise governance through defined responsibilities, institutional coordination, common evaluation principles, and accountable decision-making. The National Cyber Exercise Governance Model (C2) translates these considerations into a structured framework for national preparedness and continuous capability development.

Austria is used as the principal national reference case. The purpose is not to present Austria as a universally optimal national model, but to examine the translation from European requirements to national implementation within a concrete and sufficiently complex governance environment. Austria provides a useful reference because cyber resilience responsibilities are distributed across governmental authorities, competent bodies, sectoral structures, CSIRTs, critical entities, financial-sector governance arrangements, defence-relevant actors, research institutions, and private organizations. The Austrian case therefore allows the dissertation to examine the institutional conditions under which a coordinated national cyber exercise ecosystem would have to operate.

Section **2.2, Austria as a Reference Case for National Cyber Exercise Governance**, analyses this institutional and legal environment. It examines the Austrian strategic framework, national cybersecurity coordination architecture, the implementation of NIS2 through NISG 2026, the emerging role of the Bundesamt für Cybersicherheit, critical-entity resilience under RKEG and (Resilienz kritischer Einrichtungen-Verordnung) RKEV[68], the national and sectoral CSIRT ecosystem, financial-sector resilience and DORA-VG, Austrian AI governance, and Austria's position within the wider European cybersecurity architecture. The section also reviews cyber exercise activity in the Austrian context, including the relevance of national and multi-stakeholder

[67] Kenedli, T., & Magyar, S. (Eds.). (2025). *A mesterséges intelligencia hatásainak specifikus vizsgálata*. Katonai Nemzetbiztonsági Szolgálat. https://knbsz.gov.hu/kiadvanyok

[68] Republic of Austria. (2026). Verordnung des Bundesministers für Inneres zur Durchführung des Resilienz kritischer Einrichtungen-Gesetzes (Resilienz kritischer Einrichtungen-Verordnung – RKEV), BGBl. II Nr. 91/2026. Rechtsinformationssystem des Bundes.

exercises, and derives the principal analytical findings that inform the subsequent governance contribution.

The analysis of Austria provides the institutional context, but national comparability also requires a common descriptive language. Section **2.3, Cyber Exercise Taxonomy and Minimum Dataset**, therefore develops a formal taxonomy and information model for national cyber exercise governance. The taxonomy builds on the analytical dimensions identified in Part I but moves beyond descriptive comparison. Its purpose is to classify heterogeneous exercises consistently according to attributes such as governance, scope, capability objective, exercise format, assessed level, sector or ecosystem context, scenario and threat characteristics, recurrence, evaluation, and improvement orientation.

The Minimum Dataset complements the taxonomy by defining which core information should be recorded if exercise activities are to support national portfolio analysis and longitudinal learning. The distinction between the two is important: the taxonomy defines **how an exercise is classified**, whereas the Minimum Dataset defines **what information should be captured about it**. Together, they provide the information foundation required for national comparability without requiring identical exercise formats. Section 2.3 also establishes a layered information-governance model that distinguishes between national programme metadata, controlled programme artefacts, and restricted operational evidence. This enables programme-level visibility while preserving confidentiality, classification, data minimisation, and the need-to-know principle.

On this basis, Section **2.4, National Cyber Exercise Governance Model and Programme Blueprint**, develops the principal constructive contribution of Part II and the core contribution associated with H2. The proposed model translates the analytical findings of Part I, the Austrian reference case, and the taxonomy and dataset into a programme-level governance architecture. Its purpose is to transform heterogeneous exercise activities into a coherent national capability system.

The governance model connects national strategic priorities with programme objectives, exercise portfolio design, governance responsibilities, common artefacts, information structures, measurement, corrective-action ownership, and re-testing. It does not propose that one national authority should design or conduct every cyber exercise. Instead, it follows a federated logic in which sectoral and organizational autonomy are preserved while common national governance principles create sufficient comparability and continuity.

The model therefore shifts the focus from individual exercise completion to an evidence-oriented national improvement cycle. Its central logic is that an exercise generates resilience value only when its evidence is translated into findings, findings into corrective action, and corrective action into subsequent verification. The programme logic can therefore be expressed as:

Strategic Requirement → Programme Objective → Exercise Portfolio → Exercise → Evidence → Finding → Corrective Action → Re-test → Verified Improvement

This sequence is particularly important for the second research gap identified in the dissertation. The contribution does not assume that conducting an exercise automatically improves resilience. Instead, it establishes the governance conditions under which exercise findings can become part of a measurable and repeatable capability-development process.

Section 2.4 also incorporates the programme measurement framework. Building on the author's earlier work on the evaluation of national cyber exercise programmes, the dissertation uses a Goal-Question-Metric-oriented logic to connect national and programme-level objectives with evaluative questions, indicators, and evidence. This makes it possible to distinguish simple activity indicators, such as number of exercises or participants, from indicators that provide stronger evidence of capability development, such as remediation status, recurrence of findings, cross-sector coverage, and verified re-testing.

While the governance model defines how national exercising should be organised, Section **2.5, ENISA-Based AI-Enhanced Cyber Exercise Lifecycle**, adds the methodological process layer. The dissertation adopts the ENISA Cybersecurity Exercise Methodology as the recognised European lifecycle baseline rather than introducing an entirely independent lifecycle. The six phases -**Initiation, Design, Preparation, Execution, Evaluation, and Moving Forward** - provide the structural process through which an exercise progresses from initial purpose to follow-up improvement.

The contribution of Section 2.5 lies in defining where artificial intelligence can support this lifecycle and under which governance conditions such support should be permitted. Four cross-cutting principles guide AI integration: **human authority, traceability and provenance, bounded AI use, and phase-gated progression**. AI may support requirements analysis, objective development, scenario engineering, inject preparation, consistency checking, evidence structuring, evaluation, and reporting, but it does not replace accountable human decision-making. The governing principle is therefore that **AI may recommend and assist, while humans retain decision authority**.

The AI-enhanced lifecycle performs an important bridging function within the dissertation. The National Cyber Exercise Governance Model defines the institutional and programme environment in which exercises are governed, while the ENISA-based lifecycle defines the repeatable process through which an individual exercise is planned, executed, evaluated, and improved. Part III then takes the next step by operationalising these lifecycle-level augmentation points through the Auditable Agentic Workflow. In this way, Part II connects national governance with organizational implementation.

Section **2.6, Regulatory Pathway II – National Governance for Cyber Resilience**, develops the national segment of the dissertation's cross-cutting regulatory and policy pathway. It examines how European requirements are translated into Austrian national law, competent-authority structures, governance responsibilities, and sectoral implementation. Particular attention is given to NIS2 and NISG 2026, CER and the Austrian RKEG/RKEV framework, DORA and DORA-VG, GDPR and the Austrian data-protection framework, and the EU AI Act in the context of Austrian AI governance.

The purpose of this regulatory analysis is not to equate cyber exercises with legal compliance. An exercise cannot demonstrate compliance merely because it has been conducted. Instead, exercises are treated as **operational validation mechanisms** through which selected capabilities associated with regulatory or policy requirements can be tested. A requirement concerning continuity, incident response, third-party coordination, or human oversight can be translated into an exercise objective, observed under realistic conditions, and linked to corrective action where weaknesses are identified. This preserves the distinction between formal compliance assessment and operational evidence generation.

Part II concludes in Section **2.7, Assessment of H2 and Part II Summary**, by evaluating the national governance hypothesis against the results of the reference-case analysis and the constructive models developed in Sections 2.3–2.6. The assessment considers whether defined institutional roles, common information structures, programme-level governance, lifecycle discipline, measurement, and improvement loops provide a more coherent basis for resilience-oriented exercising than fragmented exercise activity.

The expected contribution of Part II is therefore broader than a single national programme model. It consists of a mutually reinforcing set of national-layer components: the Austrian governance reference case provides the contextual foundation; the **Cyber Exercise Taxonomy and Minimum Dataset** create the common language and information structure; the **National Cyber Exercise Governance Model and Programme Blueprint** provide the institutional and programme architecture; the **ENISA-based AI-Enhanced Cyber Exercise Lifecycle** provides the methodological process structure; and the **Regulatory Pathway II** demonstrates how European requirements are translated into national governance and exercise-relevant capabilities.

These components together form the national-level response to the central research problem. Part I established **what is converging**. Part II determines **how that emerging baseline can be governed, structured, measured, and continuously improved at national level**. Part III subsequently examines **how the AI-support elements embedded in this national and lifecycle architecture can be operationalized responsibly within organizations**.

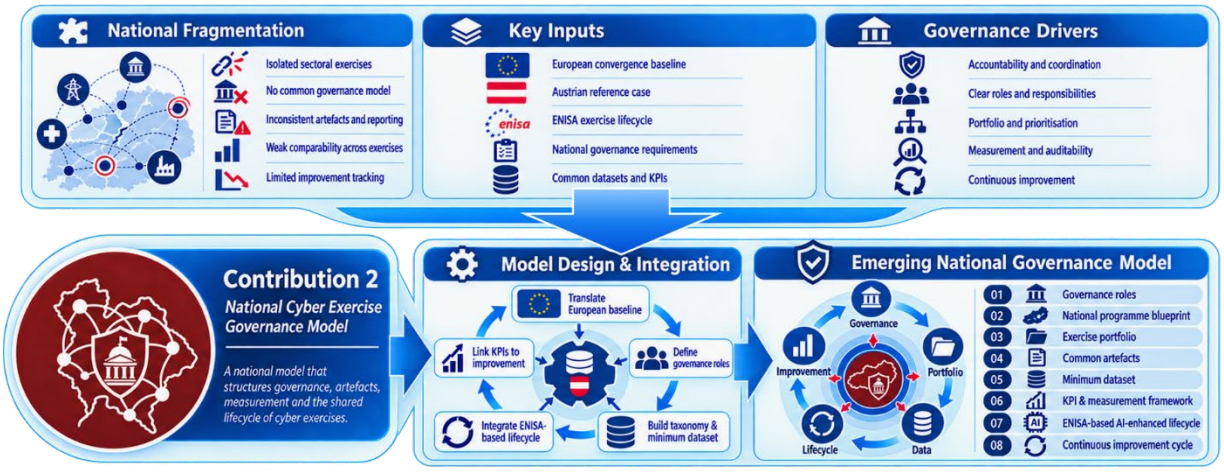


FIGURE 18 - OVERVIEW OF NATIONAL RESEARCH LAYER

Figure 18 *presents the overall logic of Part II. It illustrates how the European Convergence Framework developed in Part I is translated through the Austrian reference case into a formal Cyber Exercise Taxonomy and Minimum Dataset, the National Cyber Exercise Governance Model and Programme Blueprint, and the ENISA-based AI-Enhanced Cyber Exercise Lifecycle. The regulatory and policy pathway operates across these components, while the national governance model provides the institutional bridge between international convergence and the organizational operationalisation developed in Part III.*

2.2 Austria as a Reference Case for National Cyber Exercise Governance

Austria was selected as a purposive, information-rich reference case for examining the institutional and regulatory conditions of national cyber exercise governance. The selection was based on the coexistence of multiple governmental and operational cybersecurity responsibilities, established public-private cooperation mechanisms, recurring national cyber exercise activities, and the translation of European cybersecurity and resilience requirements into national legislation. The case study follows a qualitative, document-based approach, examining institutional mandates, strategies, legal instruments, coordination arrangements, and publicly documented exercise practices. The purpose is analytical generalisation through the identification of transferable governance requirements rather than statistical generalisation from one national environment.

Austria is used in this dissertation as the principal national reference case for examining how international and European cybersecurity and resilience requirements are translated into national governance structures and into conditions relevant for cyber exercise design. The purpose is not to present Austria as an ideal or universally transferable model. Instead, **Austria provides a realistic governance environment** in which strategic policy, civilian cybersecurity coordination, state protection, law enforcement, military cyber defence, regulatory supervision, critical-entity resilience, public-private cooperation, and operational incident response are distributed across multiple institutions.

This makes **Austria particularly suitable for H2**. The central proposition of H2 is that cyber exercises contribute more effectively to national resilience when they are embedded in a structured national capability system rather than conducted as isolated events. The Austrian case allows the dissertation to identify which actors, coordination mechanisms, legal structures, and operational interfaces would need to be connected by such a national exercise programme.
A defining characteristic of the Austrian model is that cybersecurity is treated as a whole-of-state cross-cutting responsibility. The BKA, BMI, BMEIA, BMLV and sectoral authorities, regulators, CSIRTs, critical entities, research organisations, and private operators all contribute to the national cybersecurity ecosystem. National cyber resilience therefore cannot be exercised meaningfully by testing one technical team alone. Exercises must also examine whether information reaches the correct institutions, whether escalation works, whether responsibilities are understood, whether public and private actors coordinate, and whether operational information can be translated into strategic decision-making.

Austria is used in this dissertation as the principal national reference case for examining how international and European cybersecurity and resilience requirements are translated into national governance structures, institutional responsibilities, sectoral arrangements, and cyber exercise practice. The purpose of the case is not to present Austria as a universal or optimal model of national cyber exercise governance. Rather, Austria provides a sufficiently complex and realistic governance environment in which the interaction between European regulation, national authorities, critical sectors, operational cybersecurity structures, and recurring exercise activity can be analysed.

The Austrian case is particularly relevant to the research problem because national cyber resilience is distributed across multiple institutional and sectoral actors. European requirements originating from instruments such as NIS2, the Critical Entities Resilience Directive, DORA, the EU AI Act, ENISA methodologies, and the EU Cyber Blueprint must be interpreted and implemented through national legislation, competent authorities, ministries, CSIRTs, sectoral regulators, critical entities, and public-private coordination mechanisms. This creates precisely the type of multi-actor environment in which the governance, comparability, and improvement challenges addressed by H2 become visible.

The second layer shows how these external requirements encounter the **Austrian institutional architecture**. National cybersecurity governance is distributed among several governmental, security, defence, technical-response, and regulatory actors. This institutional diversity is directly relevant to the dissertation because national cyber exercises must operate across organizational boundaries rather than within one centrally controlled authority. The Austrian case therefore demonstrates why national exercise governance requires clearly defined responsibilities, coordination mechanisms, information-sharing rules, and interfaces between strategic, operational, and technical actors.

The third layer represents the **critical sectors and stakeholders** whose resilience ultimately constitutes the operational object of national cyber preparedness. Energy, transport, finance, healthcare, water, digital infrastructure, public administration, and other essential-service environments depend on highly interconnected digital and organizational ecosystems. Their resilience cannot be assessed exclusively through isolated organizational exercises because disruptions frequently propagate through dependencies between operators, suppliers, public authorities, technical service providers, and sectoral coordination structures.

The final layer of the figure summarises the principal implications of the Austrian case for **national cyber exercise governance**. First, cyber resilience must be treated as an ecosystem capability involving people, organizations, processes, technology, and interdependencies. Second, coordination between governmental authorities, sectoral actors, private operators, and technical-response structures must itself become an object of exercising. Third, exercises require

comparable artefacts and metadata if national authorities are to understand what capabilities have been tested and where systemic gaps remain. Fourth, exercise results must feed structured improvement loops that connect findings with corrective actions and subsequent verification. Finally, cross-sector dependencies must be incorporated into exercise design if cyber exercises are to provide meaningful evidence concerning national resilience.

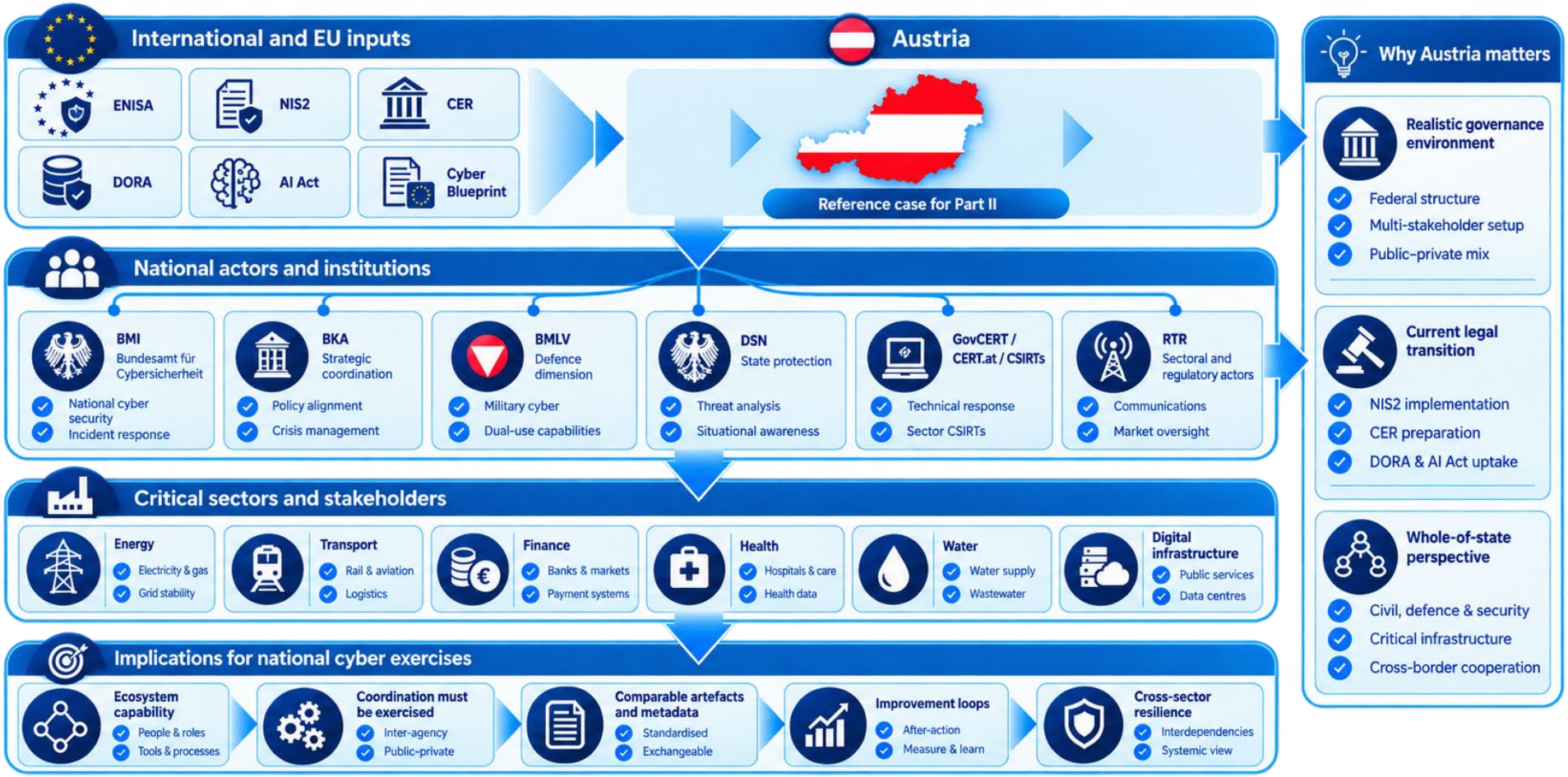


FIGURE 19 - AUSTRIA AS A REFENCE CASE FOR NATIONAL CX GOVERNANCE

The figure summarises the role of Austria as the principal national reference case in Part II. It shows how international and European inputs, including ENISA methodologies, NIS2, the Critical Entities Resilience Directive, DORA, the EU AI Act, and the EU Cyber Blueprint, are translated through Austrian institutions, cybersecurity authorities, CSIRTs, regulators, critical sectors, and public-private coordination structures. The institutional and sectoral characteristics identified in Section 2.2 provide the empirical and contextual basis for the Cyber Exercise Taxonomy and Minimum Dataset in Section 2.3 and for the National Cyber Exercise Governance Model and Programme Blueprint in Section 2.4. The governance requirements derived from the Austrian case also inform the ENISA-based AI-Enhanced Cyber Exercise Lifecycle in Section 2.5 and the national regulatory pathway analysed in Section 2.6. In this way, the figure connects the European convergence baseline developed in Part I with the national governance and lifecycle contributions of Part I

Austrian Institutional and Strategic Framework

Austria's national cybersecurity governance is embedded in a broader strategic framework that combines cybersecurity, national security, resilience, digitalisation, and increasingly artificial intelligence governance. The principal cybersecurity policy document is the **Austrian Strategy for Cybersecurity 2021 (ÖSCS 2021)**[69], which continues to provide the strategic framework for national cybersecurity policy. The strategy follows a whole-of-state and whole-of-society approach and addresses areas including governance, capability development, public-private cooperation, awareness, research, international cooperation, and resilience. Its continuing implementation is reflected in the accompanying national measures and progress reporting. including the Measures Catalogue and Progress Report 1/2025 published by the Federal Ministry of the Interior[70].

The cybersecurity strategy should be considered together with the broader **Austrian Security Strategy 2024**, which places cybersecurity, critical-infrastructure dependencies, hybrid threats, technological developments, and resilience within the wider national-security environment. This broader security perspective is particularly relevant for cyber exercise governance because major cyber incidents may affect not only information systems but also the continuity of essential state functions, critical services, public confidence, and national crisis-management structures.[71]

A further strategic component is the Austrian government's **Artificial Intelligence Mission Austria 2030 (AIM AT 2030)**[72]. Introduced in 2021, AIM AT 2030 establishes Austria's strategic framework for the development and use of trustworthy artificial intelligence. Its objectives include responsible AI use based on fundamental and human rights and European values, strengthening Austria as an AI research and innovation location, and supporting economic competitiveness. The strategy also addresses the security of AI systems, AI standards, data and infrastructure, public-sector modernisation, and the wider governance conditions required for trustworthy AI.

The **AI Implementation Plan 2024**[73] further develops AIM AT 2030 and translates the strategic framework into shorter-term governmental measures. Among its horizontal priorities are dependability, resilience and security, technological sovereignty, governance, and the implementation of the emerging European AI regulatory framework. This is particularly relevant to the present dissertation because AI-supported cyber exercise planning cannot be treated solely as a technical efficiency measure; it must be embedded within national expectations concerning security, trustworthy AI, human oversight, accountability, and responsible public-sector use.

Taken together, these strategies establish a strategic environment in which national cyber exercise objectives should not be derived only from technical threat scenarios. Exercises should also reflect national priorities concerning the continuity of essential functions, crisis governance, inter-organizational coordination, strategic communication, public-private cooperation, critical-infrastructure resilience, emerging technologies, and the responsible use of AI. Cyber exercises

---

[69] Federal Ministry of the Interior of Austria. (2021). Austrian Strategy for Cybersecurity 2021 (ÖSCS 2021). Federal Ministry of the Interior. https://www.bmi.gv.at/504/files/Cyberstrategie2021.pdf

[70] Federal Ministry of the Interior of Austria. (2025). *Measures Catalogue of the Austrian Strategy for Cybersecurity 2021: Progress Report 1/2025.* Federal Ministry of the Interior. https://www.bmi.gv.at/504/oesterreichische-strategie-fuer-cybersicherheit.html

[71] Federal Chancellery of Austria. (2024). Austrian Security Strategy 2024. Federal Chancellery of Austria.https://www.bundeskanzleramt.gv.at/dam/jcr%3A941a3cb3-7320-4c31-a4d2-3307f57faad3/sicherheitsstrategie_oesterreich_2024.pdf

[72] Austrian Federal Government. (2021). *Artificial Intelligence Mission Austria 2030 (AIM AT 2030): Strategy of the Austrian Federal Government for Artificial Intelligence.* Austrian Federal Government. https://www.digitalaustria.gv.at/dam/jcr%3A44ad1b93-6358-42b8-9f65-1e74c9a39e7f/KI%20Strategie_AIM_AT_2030_UAbf.pdf

[73] Austrian Federal Government. (2024). *Artificial Intelligence Mission Austria 2030: Implementation Plan 2024.* Austrian Federal Government. https://www.digitalaustria.gv.at/eng/strategy/strategy-AI-AIM-AT-2030.html

therefore provide a mechanism through which strategic cybersecurity and resilience objectives can be translated into observable and testable preparedness.

Austria's institutional architecture is also undergoing significant consolidation. Strategic and operational cybersecurity coordination is increasingly centred within the Federal Ministry of the Interior, including the establishment of the new national cybersecurity authority under **NISG 2026**[74]. This institutional development illustrates an important governance requirement for national exercising: organizational structures may evolve, but responsibilities, information flows, escalation paths, and decision authorities must remain sufficiently clear to be exercised and evaluated. The detailed legal and institutional implications of NISG 2026 are therefore examined separately in Section 2.2.3.

National Cybersecurity Coordination Architecture

At strategic level, the Cyber Security Steering Group (Cyber Sicherheit Steuerungsgruppe – CSS) supports implementation and further development of the national cybersecurity strategy. At operational level, the Inner Circle of the Operational Coordination Structure (IKDOK) and the wider Operational Coordination Structure (OpKoord) support information exchange, national cyber situational awareness, and coordination. Under NISG 2026, OpKoord includes the IKDOK and designated CSIRTs and can be expanded to affected essential and important entities and other organizations where required.

National cyber coordination depends not only on institutional structures but also on mechanisms capable of transforming distributed technical and organizational information into shared situational awareness and decision support. Previous research by Páhi and colleagues examined this problem from the perspectives of public-private information sharing, cyber common operating pictures, cross-organizational coordination and strategic decision-making (Páhi & Skopik, 2016[75]; Leitner, Páhi & Skopik, 2017[76]; Páhi, Leitner & Skopik, 2017[77]; Leitner, Páhi & Skopik, 2024[78]).

The Cyber Crisis Management Coordination Committee (Cyberkrisenmanagement-Koordinationsausschuss – CKM-KA) connects operational cybersecurity information with national crisis-management decision-making. This creates an exercisable governance chain: technical and organizational incident information → operational coordination and national situation picture → cyber crisis management → government-level decision-making.

For this dissertation, the structure demonstrates that national cyber resilience is an information-management and decision-coordination capability as much as a technical capability. Exercises should therefore validate reporting quality, information-sharing latency, escalation thresholds, role

[74] Republic of Austria. (2025). *Bundesgesetz zur Gewährleistung eines hohen Cybersicherheitsniveaus von Netz- und Informationssystemen (Netz- und Informationssystemsicherheitsgesetz 2026 – NISG 2026), BGBl. I Nr. 94/2025.* Rechtsinformationssystem des Bundes (RIS). RIS – NISG 2026, BGBl. I Nr. 94/2025

[75] Páhi, T., & Skopik, F. (2016). A public-private-partnership model for national cyber situational awareness. International Journal on Cyber Situational Awareness, 1, 1–23.

[76] Leitner, M., Páhi, T., & Skopik, F. (2017). Situational awareness for strategic decision making on a national level. In F. Skopik (Ed.), Collaborative cyber threat intelligence (pp. 225–276). CRC Press.

[77] Páhi, T., Leitner, M., & Skopik, F. (2017). Analysis and assessment of situational awareness models for national cyber security centers. In P. Mori, S. Furnell, & O. Camp (Eds.), Proceedings of the 3rd International Conference on Information Systems Security and Privacy (ICISSP) (pp. 334–345). SciTePress.

[78] Leitner, M., Páhi, T., & Skopik, F. (2024). Operational cyber incident coordination revisited: Providing cyber situational awareness across organizations and countries. Information Security Journal: A Global Perspective, 33(5), 486–507.

clarity, cross-sector coordination, and the usability of the national situation picture for decision-makers.

NISG 2026 and the Bundesamt für Cybersicherheit

The most important recent legal development is the Netz- und Informationssystemsicherheitsgesetz 2026 (NISG 2026)[79], promulgated as BGBl. I Nr. 94/2025 on 23 December 2025. The law implements Directive (EU) 2022/2555 (NIS2)[80]. Its main substantive provisions enter into force on 1 October 2026. Because the dissertation is finalised in August 2026, the Austrian case distinguishes explicitly between the structure operating until 30 September 2026 and the legally adopted regime that becomes operational from 1 October 2026.

NISG 2026 creates the Bundesamt für Cybersicherheit as the competent national cybersecurity authority. The authority is designed as an organizationally independent body directly subordinated to the Federal Minister of the Interior and has nationwide competence. Its responsibilities include national cybersecurity strategy coordination, large-scale incident management, national cyber situation reporting, supervision of regulated entities, international representation, the national single point of contact, the national cybersecurity coordination centre, and GovCERT-related functions.
This transition is directly relevant to H2. It demonstrates increasing institutional consolidation around defined authority, national situational awareness, coordination, reporting, oversight, and cross-sector cooperation. These functions correspond closely to the governance requirements later incorporated into the National Cyber Exercise Governance Model.

NISG 2026 also expands the regulated ecosystem across eighteen sectors. The enlarged scope means that national exercise governance must accommodate organizations with widely different maturity, resources, technical architectures, regulatory environments, and exercise experience. A national model therefore requires proportionality in exercise design combined with common metadata and evidence structures for comparability.

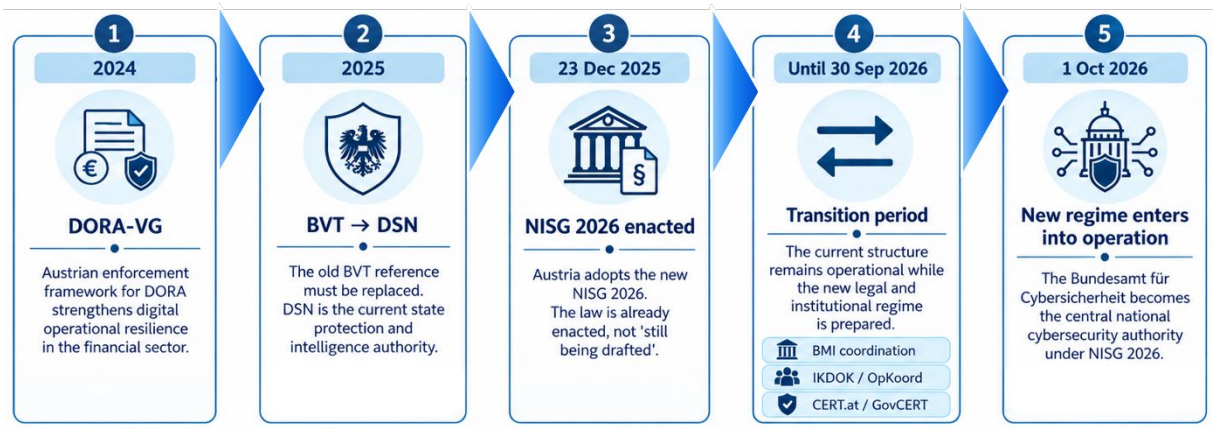


FIGURE 20 - AUSTRIA CYBERSECURITY GOVERNANCE TRANSITION

The Figure 20 illustrates Austria's cybersecurity governance transition between 2024 and 2026, highlighting key legislative and institutional developments, including DORA-VG, the transition from BVT to DSN, the adoption of NISG 2026, and the establishment of the Bundesamt für Cybersicherheit as the central cybersecurity authority. These developments strengthen national

[79] Republik Österreich. (2025). Bundesgesetz zur Gewährleistung eines hohen Cybersicherheitsniveaus von Netz- und Informationssystemen (Netz- und Informationssystemsicherheitsgesetz 2026 – NISG 2026). Bundesgesetzblatt für die Republik Österreich, BGBl. I Nr. 94/2025, Art. 1. https://www.ris.bka.gv.at/eli/bgbl/I/2025/94/20251223

[80] European Parliament & Council of the European Union. (2022). *Directive (EU) 2022/2555 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).* Official Journal of the European Union, L 333, 80–152. https://eur-lex.europa.eu/eli/dir/2022/2555/oj

coordination, clarify institutional responsibilities, and increase the importance of standardised, comparable, and auditable cyber exercises for assessing national cyber resilience.

Critical-Entity Resilience: RKEG and RKEV

Cybersecurity governance must be considered together with Austria's critical-entity resilience framework. The Resilienz kritischer Einrichtungen-Gesetz (RKEG)[81], BGBl. I Nr. 60/2025, implements the EU Critical Entities Resilience Directive and entered substantially into force on 1 March 2026. The Resilienz kritischer Einrichtungen-Verordnung (RKEV)[82], BGBl. II Nr. 91/2026, entered into force on 15 April 2026 and specifies sector-related thresholds and incident criteria.

The RKEG defines resilience broadly as the ability of critical entities to prevent, protect against, respond to, mitigate, manage, and recover from security incidents. Critical entities must conduct risk analysis and implement proportionate technical, security-related, and organizational measures documented in resilience plans. This widens the focus from cybersecurity of information systems to continuity of essential societal functions.

The exercise implication is that mature scenarios should test more than technical containment. Depending on purpose, they may need to examine essential-service continuity, alternative procedures, cross-sector dependencies, coordination with public authorities, cascading consequences, and restoration priorities. NISG 2026 and RKEG therefore reinforce the need for cross-sector and dependency-aware national exercising.

Key Actors in the Austrian Cybersecurity Ecosystem

Austria's cybersecurity ecosystem includes governmental, military, intelligence, criminal-investigation, regulatory, technical-response, critical-infrastructure, private-sector, and research actors. Their mandates differ, and this diversity creates the coordination interfaces that a national cyber exercise programme needs to validate.
The Federal Ministry of the Interior is the central civilian coordinating actor and the institutional home of the transition to the Bundesamt für Cybersicherheit. It also connects cybersecurity coordination with critical-infrastructure resilience and national crisis management. The national programme proposed in this dissertation should nevertheless be multi-institutional rather than a single-ministry exercise activity.

References in older versions of the dissertation to the Bundesamt für Verfassungsschutz und Terrorismusbekämpfung (BVT) are obsolete. The relevant current institution is the Direktion Staatsschutz und Nachrichtendienst (DSN), established in 2021 as Austria's state-protection authority and domestic intelligence service. Its remit includes terrorism, espionage, foreign intelligence activity, proliferation, and cyber threats. It contributes a national-security and intelligence dimension relevant to state-sponsored cyber operations, hybrid threats, strategic attribution, and attacks against critical infrastructure.
Law-enforcement cybercrime structures add the criminal-investigation dimension. National exercises should therefore consider the interface between incident response, evidence preservation, criminal investigation, and international police cooperation.
The Federal Ministry of Defence and Austrian Armed Forces provide the military cyber-defence dimension. Defence-oriented exercises need to test interfaces between civilian cybersecurity governance and military cyber defence while respecting their distinct constitutional mandates. Relevant objectives can include civilian-military information exchange, hybrid-threat response, national situational awareness, and coordination with European or NATO-related defence structures.

[81] Republik Österreich. (2025). Bundesgesetz zur Sicherstellung eines hohen Resilienzniveaus von kritischen Einrichtungen (Resilienz kritischer Einrichtungen-Gesetz – RKEG). Bundesgesetzblatt für die Republik Österreich, BGBl. I Nr. 60/2025, Art. 1. https://www.ris.bka.gv.at/eli/bgbl/I/2025/60/20251016

[82] Republik Österreich. (2026). Verordnung des Bundesministers für Inneres zur Durchführung des Resilienz kritischer Einrichtungen-Gesetzes (Resilienz kritischer Einrichtungen-Verordnung – RKEV). Bundesgesetzblatt für die Republik Österreich, BGBl. II Nr. 91/2026. https://www.ris.bka.gv.at/eli/bgbl/II/2026/91/20260414

National and Sectoral CSIRT Ecosystem

Austria relies on a distributed network of CSIRTs and CERTs. NISG 2026 assigns CSIRTs functions including monitoring and analysis of threats and vulnerabilities, warning and information dissemination, incident response, forensic support, vulnerability scanning, situation assessment, and participation in European cooperation.
CERT.at acts as a central national technical-response and information-sharing actor. GovCERT supports public administration and is integrated into the new NISG 2026 architecture. Sectoral and organizational response teams in areas such as energy, health, academia, telecommunications, and large enterprises provide domain-specific expertise.
The resulting governance principle is distributed technical autonomy combined with common national coordination. A national exercise programme should not centralise all response activity; it should define the minimum information, escalation, evidence, and cooperation requirements needed to create national situational awareness and collective learning.

Financial-Sector Resilience and DORA-VG

The financial sector provides an example of a highly formalised resilience-testing environment. Austria adopted the DORA-Vollzugsgesetz (DORA-VG)[83], BGBl. I Nr. 112/2024, as the national enforcement framework for Regulation (EU) 2022/2554 on digital operational resilience[84]. The framework operates alongside supervision by the Financial Market Authority and relevant cooperation with the Oesterreichische Nationalbank[85].
DORA demonstrates how testing can become part of regulated operational resilience rather than remain a voluntary training practice. A national exercise programme should not replace regulatory testing, but it can connect sectoral lessons and dependency risks to wider national crisis preparedness while respecting confidentiality and supervisory boundaries.

Artificial Intelligence Governance in Austria

Austria's AI governance environment is relevant because the dissertation later introduces AI-supported cyber exercise processes. AIM AT 2030 provides a national strategy for human-centred and trustworthy AI, while the AI Service Desk (KI-Servicestelle)[86] within Rundfunk und Telekom Regulierung (RTR)[87] has served since 2024 as a public information and support point for implementation of the EU AI Act[88].
The EU AI Act is directly applicable EU law and should not be presented as requiring an Austrian equivalent 'AI Act' in the manner of a directive transposition. The national level concerns implementation, competent authorities, coordination, supervision, guidance, and practical application. For the AI-enhanced cyber exercise lifecycle, the relevant principles are human oversight, risk-based use, transparency, data governance, security, accountability, and AI literacy.

---

[83] Republik Österreich. (2024). Bundesgesetz über das Wirksamwerden der Verordnung (EU) 2022/2554 über die digitale operationale Resilienz im Finanzsektor (DORA-Vollzugsgesetz – DORA-VG). Bundesgesetzblatt für die Republik Österreich, BGBl. I Nr. 112/2024. https://www.ris.bka.gv.at/eli/bgbl/I/2024/112/20240719

[84] European Parliament & Council of the European Union. (2022). Regulation (EU) 2022/2554 of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, (EU) No 909/2014 and (EU) 2016/1011. Official Journal of the European Union, L 333, 1–79. https://eur-lex.europa.eu/eli/reg/2022/2554/oj

[85] Financial Market Authority. (2025, December 16). BMI, FMA and OeNB cybersecurity cooperation. https://www.fma.gv.at/en/bmi-fma-and-oenb-cybersecurity-cooperation/

[86] The **AI Service Desk (KI-Servicestelle)** established within RTR in 2024 serves as a public point of contact and information hub for artificial intelligence and supports the practical implementation of the European AI Act. Its legal basis is § 20c KOG and § 194a TKG, introduced through BGBl. I No. 6/2024 (Rundfunk und Telekom Regulierungs-GmbH, 2024

[87] Rundfunk und Telekom Regulierungs-GmbH. (2024). *Service Desk for Artificial Intelligence*. RTR. RTR AI Service Desk

[88] European Parliament and Council of the European Union. (2024). *Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence 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, L 2024/1689*. EUR-Lex: EU AI Act

Austria in the European Cybersecurity Architecture

Austria is embedded in European cybersecurity cooperation through the NIS Cooperation Group, the CSIRTs Network, EU-CyCLONe, ENISA, and the European Cybersecurity Competence Community. These mechanisms represent strategic, operational, and technical levels of European cooperation and mirror the layered nature of Austria's national architecture.
National exercises should therefore include European escalation and cooperation dimensions where relevant, including cross-border notification, technical CSIRT cooperation, operational crisis coordination, shared situational awareness, and alignment with EU crisis-management mechanisms.

Cyber Exercises in the Austrian Context

Austria already possesses relevant exercise experience. Recurring KSÖ Cybersecurity Planspiele, cooperation with AIT and cyber range capabilities, public-private participation, and international involvement demonstrate that several practical preconditions for a national exercise ecosystem already exist.
The proposed National Cyber Exercise Governance Model therefore does not claim to invent cyber exercising in Austria. Its contribution is the programme layer through which existing and future exercises can become more comparable, measurable, evidence-oriented, and connected across sectors and exercise cycles.

Analytical Findings from the Austrian Reference Case

Eight findings from the Austrian reference case directly inform the design of the national contribution:

1. National cyber resilience is an ecosystem capability; no single institution possesses all relevant responsibilities, data, technical capabilities, and decision authority.
2. Coordination itself must be exercised; strategic steering, operational coordination, CSIRT cooperation, and crisis-management escalation require practical validation.
3. National situational awareness is a measurable capability; timeliness, completeness, accuracy, relevance, escalation speed, and decision usefulness can be observed during exercises.
4. Civilian and military cyber structures should remain distinct but interoperable; exercises should test interfaces rather than erase constitutional and organizational boundaries.
5. Critical-infrastructure resilience requires cross-sector exercising because cyber incidents can create physical, societal, economic, and continuity consequences across dependencies.
6. Regulatory diversity requires a common exercise abstraction. NISG 2026[89], (Resilienz kritischer Einrichtungen-Gesetz) RKEG[90], DORA, GDPR, and AI governance differ in purpose, but their operational requirements can be mapped to common objectives, artefacts, evidence, and improvement processes.
7. Confidentiality limits centralised measurement. National comparability should rely on standardised metadata and aggregated findings while sensitive raw evidence remains distributed and access-controlled.
8. Exercises can function as governance laboratories by revealing unclear roles, unrealistic assumptions, coordination bottlenecks, and gaps between formal procedures and actual crisis performance.

Austria as the Reference Point for the National Governance Contribution

The Austrian case demonstrates that the essential actors and legal mechanisms of a national cybersecurity ecosystem can exist while programme-level structures for systematic exercise

[89] Republic of Austria. (2025). *Bundesgesetz zur Gewährleistung eines hohen Cybersicherheitsniveaus von Netz- und Informationssystemen (Netz- und Informationssystemsicherheitsgesetz 2026 – NISG 2026), BGBl. I Nr. 94/2025.* Rechtsinformationssystem des Bundes. Official NISG 2026 text in RIS

[90] Republic of Austria. (2025). *Bundesgesetz zur Sicherstellung eines hohen Resilienzniveaus von kritischen Einrichtungen (Resilienz kritischer Einrichtungen-Gesetz – RKEG), BGBl. I Nr. 60/2025.* Rechtsinformationssystem des Bundes. Official RKEG text in RIS

comparability and longitudinal improvement remain incomplete. This observation provides the constructive starting point for the National Cyber Exercise Governance Model.
The proposed model builds on rather than replaces Austrian structures. It adds a programme layer connecting national strategic priorities, regulatory requirements, competent authorities, sectoral stakeholders, exercise portfolios, taxonomy, common artefacts, minimum datasets, KPIs, After-Action Reports, corrective actions, and re-testing.
The transition can be summarised as: European convergence and regulation → Austrian law and governance → distributed authorities, CSIRTs, critical entities and exercise actors → need for common exercise governance → National Cyber Exercise Governance Model → comparable evidence and continuous resilience improvement.

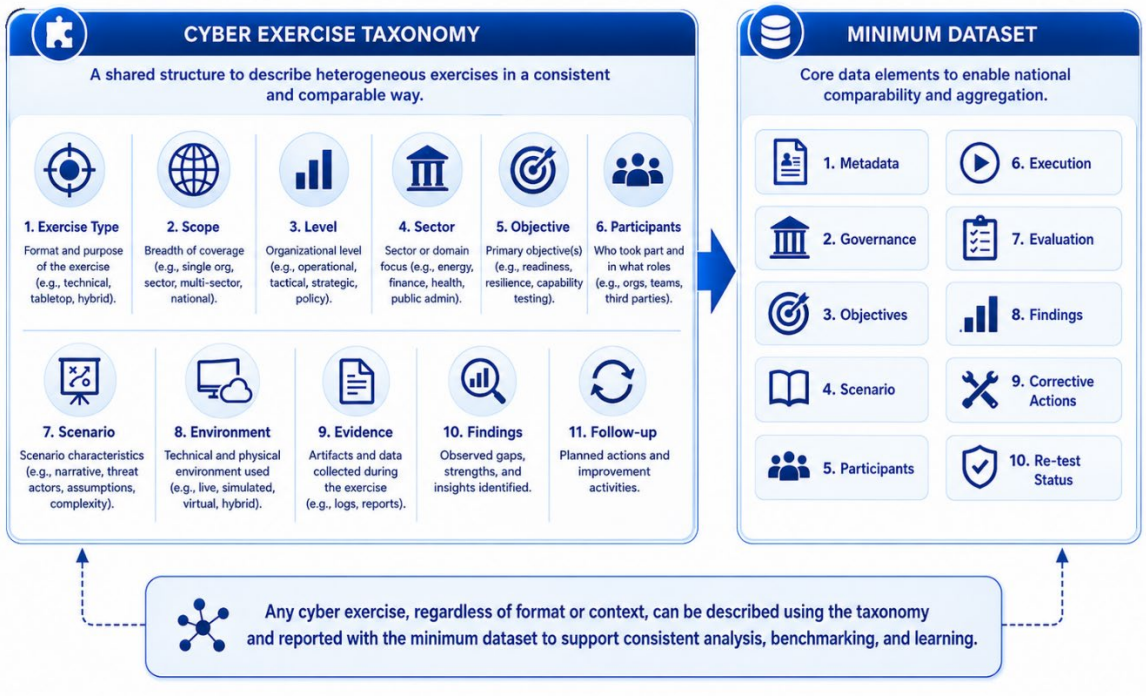


FIGURE 21 - CX TAXONOMY AND MINIMUM DATASET

A key data-governance principle follows: sensitive exercise evidence may remain distributed within the originating organization or competent authority, while standardised metadata and approved findings can be aggregated at programme level. The model therefore seeks comparability without requiring unrestricted centralisation of operationally sensitive information.
The corresponding minimum dataset operationalises this taxonomy for national programme use. Core fields include metadata, governance, objectives, scenario, participants, execution information, evaluation, findings, corrective actions, and re-test status. The minimum dataset is deliberately smaller than the full exercise documentation set: its purpose is national comparability and longitudinal analysis, not central storage of all sensitive exercise artefacts.
The proposed taxonomy describes exercises through common dimensions including identity and organiser, attribution level, capability objectives, sector, exercise type, assessed level, scenario theme, environment, participants, duration, cadence, evidence, findings, follow-up, and re-test status. These dimensions allow a strategic tabletop, a technical range exercise, a sectoral functional exercise, and a multinational crisis exercise to remain operationally different while still being represented through a common descriptive structure.
The international analysis in Part I demonstrates that the label 'cyber exercise' covers substantially different formats. For national governance, this diversity becomes a data and comparability problem. The purpose of the national taxonomy is not to standardise scenarios or force all exercises into one operational format; it is to standardise the metadata required to understand what was tested, under which conditions, by whom, and with what results.

2.3 Cyber Exercise Taxonomy and Minimum Dataset

The international and European analysis in Part I demonstrated that cyber exercises are inherently heterogeneous. They differ in purpose, scope, governance, technical depth, participant composition, sectoral context, frequency, and evaluation approach. This diversity is not in itself a weakness. Different exercise objectives require different formats, levels of realism, stakeholder groups, and technical environments. A strategic national crisis-management exercise, for example, cannot be designed in the same way as a live-fire technical exercise or an intelligence-led adversarial test. The comparability problem arises not because these exercises are different, but because the information used to describe, evaluate, and govern them is often inconsistent.

For national cyber exercise governance, this creates a practical limitation. If exercises are described using different terminology, incompatible documentation structures, and inconsistent evaluation criteria, national stakeholders cannot reliably determine which capabilities have been tested, which sectors remain under-exercised, which weaknesses recur across exercise cycles, or whether previously identified findings have been addressed. A national exercise programme therefore requires a common descriptive and data structure through which heterogeneous exercises can be represented without requiring them to become methodologically identical.

This chapter addresses that requirement through the development of a **Cyber Exercise Taxonomy and Minimum Dataset**. The taxonomy provides a common classification logic for describing the essential characteristics of a cyber exercise, while the minimum dataset defines the core information that should be recorded if exercises are to support national portfolio management, comparability, evidence generation, and longitudinal improvement tracking.

The distinction between the two components is important. The **taxonomy defines how an exercise is classified**, while the **minimum dataset defines what information must be recorded about it**. Together, they create a structured information layer that connects individual exercise activities with the broader National Cyber Exercise Governance Model developed in Section 2.4.

The proposed model does not seek to standardise all cyber exercises into a single format. Instead, it standardises the metadata and evidence-related information required to understand what each exercise was intended to test, how it was conducted, which stakeholders participated, what evidence was generated, which findings resulted, and whether corrective actions were subsequently implemented and re-tested.

In this way, the taxonomy and minimum dataset respond directly to the comparability and improvement gaps identified in the introductory research framework. They provide the information structure necessary to transform individual exercise events into elements of a cumulative national capability-development system.

Derivation and Design Logic of the Taxonomy

The taxonomy is derived from the cumulative research logic of the dissertation rather than from a single existing exercise methodology. Its construction draws on three analytical foundations.

The first foundation is the comparative analysis of international and national cyber exercise reference cases presented in Part I. That analysis demonstrated that exercises can be meaningfully compared when a common set of analytical dimensions is applied, including governance, scope, exercise objective, format, assessed capability level, sectoral context, threat scenario, recurrence, and evaluation approach. These dimensions provide the empirical basis for determining which characteristics are sufficiently stable and relevant to support national-level classification.

The second foundation is the comparison of established exercise methodologies and standards. HSEEP, ENISA, BSI, and related guidance differ in terminology and process structure, but repeatedly emphasize common elements such as defined objectives, clear roles, lifecycle orientation, scenario structure, exercise artefacts, evaluation planning, evidence collection, after-

action reporting, corrective actions, and continuous improvement. These recurring characteristics indicate that national comparability must extend beyond basic descriptive metadata and include information about the quality and evaluation structure of an exercise.

The third foundation is the national governance requirement examined through the Austrian reference case. A national exercise system involves organizations with different legal responsibilities, sectoral backgrounds, levels of maturity, technical capabilities, and confidentiality requirements. National governance therefore requires a model that provides enough common structure for steering and learning while allowing individual organizations to retain control of sensitive operational information.

The taxonomy is consequently designed according to four principles.

First, it must support **comparability without uniformity**. The model should enable a technical live-fire exercise, an executive tabletop exercise, and a cross-sector national exercise to be represented through a common structure while retaining the characteristics that make each exercise distinct.

Second, it must support **portfolio visibility**. National stakeholders should be able to determine which sectors, capability levels, exercise formats, and strategic objectives are covered by the national exercise portfolio and where gaps remain.

Third, it must support **longitudinal analysis**. Exercise data should be reusable across cycles so that recurring findings, repeated capability weaknesses, and evidence of improvement can be tracked over time.

Fourth, it must support **governance under confidentiality constraints**. National comparability cannot depend on the unrestricted centralization of sensitive technical or organizational information. The model must therefore distinguish between programme-level metadata and restricted exercise evidence.

These principles lead to a two-level structure. The taxonomy defines the conceptual classification categories, while the minimum dataset translates those categories into concrete data fields suitable for national programme governance.

The resulting logic can be expressed as follows:

International exercise diversity → recurring analytical dimensions → formal taxonomy → minimum dataset → national portfolio visibility → evidence aggregation → corrective action tracking → re-testing.

The taxonomy should therefore be understood as a governance-enabling instrument rather than merely a classificatory tool. Its purpose is to make otherwise heterogeneous cyber exercises manageable as part of a coherent national capability system.

Cyber Exercise Taxonomy

The proposed Cyber Exercise Taxonomy contains a set of complementary dimensions through which individual exercises can be classified consistently. These dimensions are intended to be sufficiently general to support national comparability while remaining flexible enough to accommodate sector-specific, defence-oriented, and classified exercise environments.

The first dimension concerns **exercise identity and governance**. Every exercise should be identifiable as a distinct activity or programme cycle. Relevant attributes include the official exercise name, edition or year, lead organiser, co-organizers, programme ownership, and responsible authority where applicable. These attributes establish institutional accountability and allow recurring exercise series to be linked across cycles.

The second dimension concerns **scope and attribution level**. For the purposes of national governance, four principal levels are distinguished: organizational, sectoral, national, and

international or multinational. These levels describe the community being exercised and the governance environment in which the exercise operates. Organizational exercises primarily validate internal processes and capabilities. Sectoral exercises test coordination and dependencies among organizations delivering related services. National exercises address cross-sector coordination, public-private cooperation, national crisis-management structures, and whole-of-government or whole-of-society preparedness. International exercises additionally test interoperability, cross-border coordination, multinational command structures, and shared situational awareness.

The third dimension concerns the **primary capability objective**. This dimension identifies what the exercise is intended to test or improve. Relevant capability areas may include incident detection and response, crisis management, executive decision-making, situational awareness, information sharing, strategic communication, continuity and recovery, third-party coordination, regulatory response, cyber defence operations, military cyber operations, or cross-sector coordination. A single exercise may address several capability objectives, but one or more primary objectives should be identified explicitly.

The fourth dimension concerns the **exercise format**. The taxonomy distinguishes between discussion-based or tabletop exercises, functional exercises, distributed exercises, technical or live-fire exercises, red-team or blue-team exercises, threat-led or adversarial testing, full-scale exercises, and hybrid formats. These categories reflect the principal execution model through which participant performance is stimulated and observed.

The fifth dimension concerns the **assessed capability level**. Four broad levels are distinguished: technical, tactical or operational, strategic or executive, and multi-level. This dimension is necessary because meaningful comparison requires clarity regarding what type of performance is being assessed. A technical measure such as detection latency, for example, cannot be interpreted in the same way as a strategic decision-making metric.

The sixth dimension concerns the **sectoral and ecosystem context**. Exercises may focus on a specific critical sector, multiple interdependent sectors, public administration, defence, essential services, or a broader whole-of-society environment. Where useful, sector classifications such as those reflected in European resilience regulation may provide a common coding structure. The taxonomy should nevertheless allow categories such as defence or military operations that do not map naturally to civilian sector frameworks.

The seventh dimension concerns the **scenario and threat context**. Relevant attributes include the principal threat or disruption type, threat actor where applicable, affected functions, key dependencies, and the broader crisis context. Scenarios may include ransomware, supply-chain compromise, destructive malware, espionage, hybrid threats, operational technology disruption, denial of service, insider activity, state-sponsored operations, or compound crises. Recording this context is important because scenario complexity and threat realism influence the interpretation of exercise outcomes.

The eighth dimension concerns **scale and temporal characteristics**. Relevant attributes include exercise duration, participant or organization count, frequency, cadence, and relationship to previous or future exercise cycles. These attributes allow national stakeholders to distinguish between isolated events and sustained exercise programmes and to analyse resource intensity and recurrence.

The ninth dimension concerns **evaluation and evidence orientation**. This dimension captures whether formal exercise objectives are evaluated, whether evidence is collected systematically, which evaluation methodology is used, whether an After-Action Report or equivalent output is produced, and whether findings are linked to corrective actions. This classification is particularly important for distinguishing training-oriented activities from evidence-oriented exercises embedded in an improvement cycle.

The tenth dimension concerns **improvement and re-testing**. Exercises should be classified according to whether identified findings are assigned to responsible owners, whether corrective actions are tracked, and whether re-testing is planned or completed. This dimension is central to the dissertation's resilience-oriented perspective because national value depends not only on identifying weaknesses, but on demonstrating whether they are addressed over time.

Taken together, these dimensions create a multidimensional representation of cyber exercises. They allow legitimate differences in exercise design to remain visible while introducing sufficient structure for systematic comparison and governance.

Taxonomy dimension	Purpose	Example values
Exercise identity	Uniquely identifies the exercise	Name, edition, year, exercise ID
Governance	Identifies institutional ownership	Authority, ministry, operator, military organization, EU body, partnership
Scope	Defines exercised community	Organizational, sectoral, national, international
Capability objective	Defines what is being tested	Incident response, crisis management, continuity, coordination, communication
Exercise format	Defines execution model	TTX, functional, distributed, live-fire, red/blue, hybrid
Assessed level	Defines performance layer	Technical, operational, strategic, multi-level
Sector / ecosystem	Defines operational context	Energy, health, finance, transport, public administration, defence, cross-sector
Scenario / threat	Defines disruption context	Ransomware, supply-chain, APT, OT disruption, hybrid threat
Scale / recurrence	Defines implementation characteristics	Duration, participants, annual, biennial, ad hoc
Evaluation / evidence	Defines evidence-generation structure	Observation, scoring, AAR, evaluation report
Improvement	Defines follow-up structure	Corrective action, owner, due date, re-test

TABLE 9 - PROPOSED CX TAXONOMY

The taxonomy is deliberately broader than a traditional exercise-type classification. Its purpose is not merely to distinguish tabletop exercises from technical exercises, but to describe the full governance and evidence context in which an exercise operates.

Minimum Dataset for National Comparability

The taxonomy provides the conceptual structure for classification, but national programme governance requires a concrete and reusable data model. The dissertation therefore proposes a **Minimum Cyber Exercise Dataset** containing the information that should be recorded for each exercise if it is to contribute to portfolio management, comparability, evidence aggregation, and longitudinal improvement tracking.

The use of a structured minimum dataset also builds on earlier research into the systematic identification and integration of heterogeneous information sources for national cyber situational awareness (Páhi et al., 2018[91]). While the earlier work focused on the information required to construct cyber common operating pictures, the present dissertation transfers the underlying principle of structured and comparable data collection to national cyber exercise governance.

[91] Páhi, T., Skopik, F., Kieseberg, P., & Leitner, M. (2018). Erhebung von Informations- und Datenquellen für Cyber-Lagebilder. In F. Skopik, T. Páhi, & M. Leitner (Eds.), Cyber situational awareness in public-private-partnerships: Organisationsübergreifende Cyber-Sicherheitsvorfälle effektiv bewältigen (pp. 191–236). Springer Vieweg.

The minimum dataset is designed around the principle that national governance requires enough information to understand the exercise and its outcomes, but does not require unrestricted access to all underlying sensitive or classified material.

The first group of fields concerns **identification**. Each exercise should have a unique exercise identifier, official exercise name, edition or year, and date. These fields enable longitudinal linking and avoid ambiguity between recurring exercise series and individual iterations.

The second group concerns **governance and ownership**. Relevant fields include lead organiser, supporting organizers, responsible programme, responsible authority, and where appropriate, exercise sponsor. This enables national governance bodies to identify responsibility and institutional ownership.

The third group concerns **scope and participation**. Fields should include attribution level, participating organizations, sector or ecosystem context, participant numbers, and relevant external or international participants. This allows exercise coverage to be analysed at programme level.

The fourth group concerns **exercise purpose and design**. Minimum fields include exercise objectives, capability areas tested, exercise format, assessed level, scenario theme, principal threat or disruption type, and key dependencies represented. These fields enable exercises to be compared according to purpose rather than title alone.

The fifth group concerns **execution characteristics**. Relevant fields include duration, location or delivery model where appropriate, technical environment if relevant, and whether the exercise is part of a recurring programme.

The sixth group concerns **evaluation and evidence**. At minimum, the dataset should identify the evaluation method, evidence sources, principal exercise artefacts, and whether a formal After-Action Report or equivalent report was produced. The purpose is not to centralise all exercise evidence, but to establish whether evidence exists and how it was generated.

The seventh group concerns **findings and improvement**. Minimum fields should record key findings or finding categories, priority or severity where appropriate, responsible corrective-action owner, target completion date, current status, whether re-testing is required, and whether re-testing has occurred.

The eighth group concerns **governance constraints**. Relevant fields include information classification or sensitivity, legal or regulatory relevance, and data-sharing restrictions. This is particularly important in defence and critical-infrastructure environments, where exercise data may contain sensitive technical information, vulnerabilities, organizational weaknesses, or classified operational details.

The proposed minimum dataset is summarised in Table 9 :

Data group	Minimum field
Identification	Exercise ID
Identification	Exercise name
Identification	Edition / year
Identification	Exercise date

Governance	Lead organiser
Governance	Co-organisers / supporting partners
Governance	Responsible programme / authority
Scope	Organizational / sectoral / national / international
Scope	Participating organizations
Scope	Sector / ecosystem
Scope	Participant scale
Purpose	Exercise objectives
Purpose	Capability areas tested
Design	Exercise format
Design	Assessed level
Design	Scenario theme
Design	Threat / disruption type
Design	Critical dependencies represented
Execution	Duration
Execution	Delivery / technical environment where relevant
Evidence	Evaluation method
Evidence	Evidence sources
Evidence	Main artefacts produced
Evidence	AAR / evaluation report available
Results	Main findings / finding categories
Results	Priority / severity
Improvement	Corrective-action owner

Improvement	Target completion date
Improvement	Current status
Improvement	Re-test required
Improvement	Re-test date / linked exercise
Governance	Sensitivity / classification
Regulatory	Relevant legal / regulatory / policy requirement
Regulatory	Data-sharing restrictions

TABLE 10 - MINIMUM CX DATASET

The minimum dataset should not be interpreted as requiring a national authority to centralise all sensitive exercise evidence. Such an approach would be unrealistic in defence-oriented, regulated, or critical-infrastructure environments and could itself create security and confidentiality risks.

The dissertation therefore distinguishes between three levels of information.

The first level is **national programme metadata**. This includes the minimum information required to understand what was tested, which organizations or sectors participated, which capability areas were covered, and whether significant findings and improvement actions exist.

The second level is **controlled programme artefacts**. These may include exercise concepts, objective mappings, evaluation summaries, AARs, improvement plans, or anonymised evidence that can be shared with designated national authorities under defined access conditions.

The third level is **restricted operational evidence**. This may include detailed technical logs, vulnerabilities, classified scenarios, sensitive organizational weaknesses, or attack artefacts that remain under the control of the participating entity or competent authority.

This layered approach allows national governance to obtain sufficient visibility for steering and improvement without violating confidentiality, national security, commercial sensitivity, or data-protection requirements.

The minimum dataset therefore follows a **data-minimisation and need-to-know principle**. National comparability does not require that all organizations disclose all underlying evidence. It requires a sufficiently consistent record of what was tested, what was observed, which improvement action resulted, and whether that action was subsequently verified.

Integration into National Exercise Governance

The Cyber Exercise Taxonomy and Minimum Dataset developed in the previous sections are intended to function as integral components of the broader National Cyber Exercise Governance Model rather than as isolated classification instruments. Their primary purpose is to create the informational conditions required for coherent programme steering, cross-organizational comparability, longitudinal analysis, and evidence-based improvement. In this sense, the taxonomy provides the common conceptual language through which different exercise formats can be described, while the minimum dataset ensures that the information necessary for governance is captured in a sufficiently consistent manner across organizations, sectors, and exercise cycles.

At national level, this common information structure is important because cyber exercise activities are typically distributed across multiple institutions with different mandates, maturity levels, sectoral responsibilities, and security requirements. Without a common descriptive basis, exercise portfolios remain difficult to oversee. National authorities may know that exercises are being conducted, but they may lack sufficient visibility into which capabilities are being tested, which sectors are covered, whether exercise activity is balanced across strategic, operational, and technical levels, or whether recurring weaknesses are being addressed over time. The taxonomy and minimum dataset therefore make it possible to move from an event-oriented perspective towards a programme-oriented understanding of cyber exercising.

One of the principal governance functions enabled by the taxonomy is portfolio visibility. When exercises are classified according to common dimensions, national stakeholders can analyse the overall distribution of exercise activity and identify structural gaps. This may reveal, for example, that a national portfolio contains a large number of technically oriented exercises but relatively few strategic crisis-management exercises, or that certain critical sectors participate regularly while others remain underrepresented. It may also reveal that particular capability areas, such as third-party coordination, continuity management, strategic communication, or cross-sector information sharing, are not being tested with sufficient frequency. The taxonomy therefore supports a more deliberate allocation of exercise activity and allows national programme managers to align exercise portfolios more closely with identified resilience priorities.

A second governance function concerns capability progression. Organizations and sectors differ significantly in their level of preparedness and in the complexity of exercises they are able to conduct effectively. A national programme should therefore not require all participants to perform the same exercise type at the same level of maturity. Instead, the taxonomy can support a staged progression model in which organizations move from less complex formats, such as plan reviews, communication drills, and tabletop exercises, towards functional, technically supported, multi-entity, or live-fire exercises as their capabilities develop. In this way, national governance can preserve flexibility while still promoting systematic capability development.

The minimum dataset complements this portfolio perspective by supporting longitudinal analysis across exercise cycles. When core data fields are recorded consistently, the results of individual exercises can be connected over time. This enables national stakeholders to identify recurring weakness categories, monitor corrective actions, and determine whether previously identified deficiencies have subsequently been re-tested. The value of the dataset therefore lies not only in documenting what happened during one exercise, but in creating a cumulative evidence base that can support resilience-oriented improvement.

This longitudinal perspective is particularly important in relation to the second research gap identified in the dissertation. The conduct of an exercise alone does not demonstrate that resilience has improved. Improvement becomes observable only when exercise findings are translated into corrective actions, assigned to responsible owners, monitored, and subsequently verified. The taxonomy and dataset provide the structural basis for linking these stages. In this context, the national governance logic can be described as a continuous sequence in which exercise classification and portfolio visibility lead to evidence generation, findings, corrective actions, re-testing, and verified improvement.

The same structure also enables a distinction between organization-specific findings and systemic national-level weaknesses. Individual exercise results may contain sensitive technical or operational information that cannot be shared broadly. Nevertheless, recurring categories of weakness may still be identified at an aggregated level. If several exercises reveal problems in escalation procedures, dependency management, supplier coordination, crisis communication, recovery prioritisation, or cross-organizational information sharing, these patterns may indicate a broader governance issue rather than an isolated organizational deficiency. The minimum dataset

therefore allows national programme governance to identify cross-cutting resilience problems without requiring unrestricted access to all underlying operational evidence.

This capability is particularly relevant for national coordination because systemic findings may require responses beyond the level of a single organization. Depending on the nature of the weakness, appropriate measures may include revised national guidance, new coordination procedures, targeted sectoral exercises, common templates, additional training, clarification of institutional responsibilities, or focused re-testing. The taxonomy and dataset therefore support not only observation but also prioritisation and intervention at programme level.

A further function concerns the connection between exercise governance and the regulatory and policy pathway developed throughout the dissertation. International and European requirements relating to cybersecurity risk management, continuity, incident response, resilience, third-party dependencies, data governance, and accountability can be translated into exercise objectives and capability areas. Through the minimum dataset, it becomes possible to document whether a requirement-relevant capability has been exercised, what type of evidence was generated, which weaknesses were identified, and whether corrective actions have subsequently been implemented.

This relationship must, however, be interpreted carefully. The completion of a cyber exercise does not constitute proof of legal compliance. Exercises can provide operational evidence concerning whether specific capabilities or procedures function under realistic conditions, but formal compliance remains dependent on the applicable legal framework, competent authority, and the wider body of organizational controls and documentation. The role of exercises is therefore better understood as operational validation rather than certification.

Accordingly, the relevant logic is not that exercise completion demonstrates compliance, but that regulatory and policy requirements can be translated into exercise objectives, which in turn generate observable evidence, findings, corrective actions, and subsequent re-testing. This approach strengthens the practical value of exercises within governance and assurance processes while avoiding an overstatement of their legal significance.

The integration of the taxonomy and minimum dataset into national governance also has implications for information management and confidentiality. In defence, critical-infrastructure, and regulated environments, detailed exercise information may contain classified, commercially sensitive, or security-relevant material. A national programme should therefore not assume that all underlying evidence can or should be centrally stored. Instead, the governance model should distinguish between programme-level metadata, controlled artefacts, and restricted operational evidence. This allows national authorities to obtain sufficient visibility for portfolio management and improvement tracking while respecting the need-to-know principle, data minimisation, national security requirements, and organizational confidentiality.

The proposed information model also creates an important methodological bridge to the ENISA-based AI-enhanced cyber exercise lifecycle developed in Section 2.5. Structured exercise data can serve as controlled inputs to lifecycle activities and make the relationships between requirements, objectives, scenarios, injects, evidence, findings, and corrective actions more explicit. At the same time, information-classification fields and access restrictions can define which data may be used by AI-supported functions and which data must remain excluded from automated processing. The taxonomy therefore supports not only comparability but also more controlled and traceable AI integration.Within the broader architecture of the dissertation, the relationship between these elements is sequential and complementary. The taxonomy establishes a common descriptive language, the minimum dataset creates a comparable information structure, the National Cyber Exercise Governance Model defines institutional responsibilities and programme steering, the ENISA-based lifecycle provides the repeatable process structure, and the Auditable Agentic Workflow later operationalises AI support within that governed environment. None of these elements is intended to function independently; together, they form a coherent national exercise ecosystem.

The Cyber Exercise Taxonomy and Minimum Dataset should therefore be understood as enabling components of Contribution 2. Their scientific value lies in translating the comparability requirement identified at the international and European level into a practical national governance mechanism. They make it possible to retain legitimate diversity in exercise design while still creating the conditions for portfolio visibility, evidence aggregation, corrective-action tracking, and longitudinal assessment.

The central principle is that national comparability does not require identical cyber exercises. It requires **consistent information about different cyber exercises**. By standardising how exercises are described, classified, and connected to evidence and improvement, the proposed taxonomy and minimum dataset provide the informational foundation required for a coherent, scalable, and evidence-oriented national cyber exercise governance system.

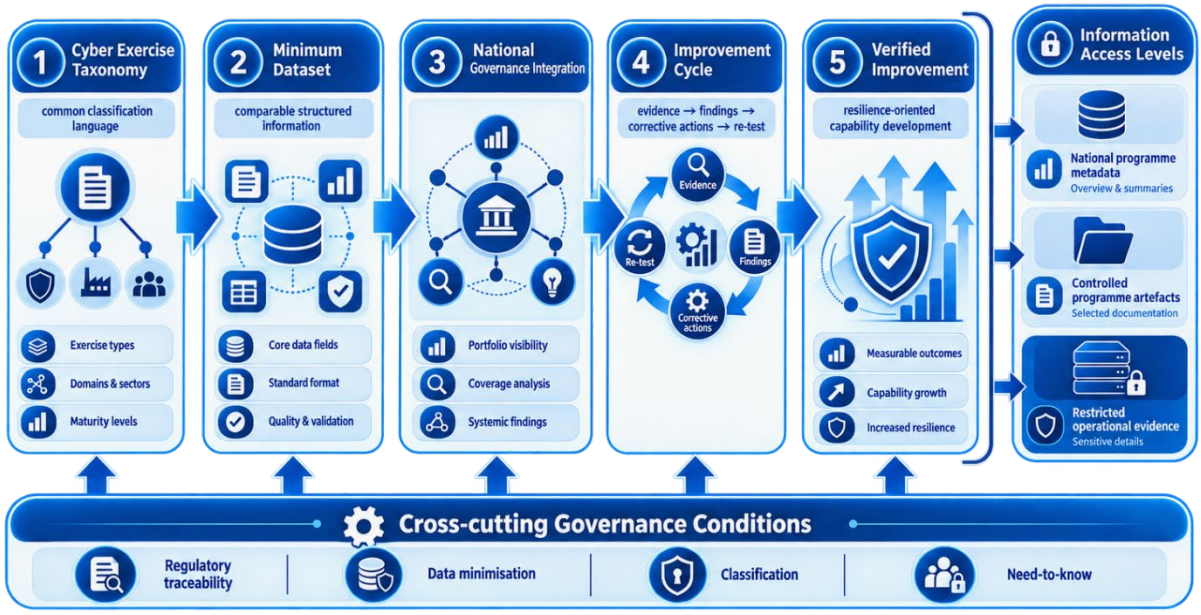


TABLE 11 - INTEGRATION OF THE CX TAXONOMY INTO NATIONAL EXERCISE GOVERNANCE

The figure illustrates how the Cyber Exercise Taxonomy establishes a common classification language, while the Minimum Dataset provides a comparable information structure for national programme governance. Together, they enable portfolio visibility, evidence aggregation, systemic gap identification, corrective-action tracking, and re-testing, while differentiated information-access levels preserve confidentiality and data minimization.

2.4 National Cyber Exercise Governance Model (C2) and Programme Blueprint

*A **national cyber exercise programme** is a governance and capability-development instrument designed to turn dispersed, heterogeneous exercise activities into measurable and sustained resilience improvement at state level.* Many countries already host a wide spectrum of exercise formats, internal tabletops, sector drills, crisis-management staff exercises, range-based technical events, red–blue simulations, and participation in international flagship exercises.

These activities can be valuable for the organizations involved, yet they do not automatically produce a national effect. ***Without a programme layer, exercises remain difficult to compare, lessons learned are not aggregated, and recurring systemic gaps persist across sectors***

and years. *A national programme therefore addresses a different problem than how to run a good exercise. Its focus is to*

- *align exercises with national resilience priorities,*
- *standardise a minimum set of artefacts and metadata to enable comparability,*
- *institutionalise after-action improvement loops and follow-up verification, and*
- *provide enabling infrastructure that lowers the barrier for organizations with limited planning maturity while preserving flexibility for advanced participants.*

At the conceptual core is the distinction between exercises that are run under programme rules and those that are not. In practice, a national authority cannot and should not attempt to control all exercise activities. Sector communities, multinational companies, and security providers will continue to run exercises that follow their own standards and confidentiality constraints. A programme that aims to be effective must therefore be inclusive by design: it defines what qualifies as a "programme exercise" (i.e., meets the standardised planning and reporting expectations) while still providing a path for "non-programme exercises" to contribute to national learning through a confidentiality-aware minimum dataset. This dual-track architecture is essential for scaling: it incentivises standardisation without stifling existing initiatives and allows national learning to draw on a broader evidence base than the programme itself could directly generate.

National Cyber Strategies – Zero Point

National cyber exercise programmes should not be designed as standalone training initiatives. *They are most effective when they are explicitly embedded in a broader national cyber strategy that defines the state's cyber resilience ambition, prioritises critical functions, and establishes a governance logic for preparedness, response, and recovery.* A national cyber strategy provides the "why" and the "what": it articulates national objectives (e.g., continuity of essential services, protection of critical infrastructure, societal trust, and national security), identifies priority risks and dependencies, and allocates responsibilities across government, regulators, operators, and supporting ecosystems. Cyber exercises, in turn, provide the "how" for preparedness validation: they translate strategic intent into observable performance under stress, reveal gaps that cannot be detected through policy review alone, and generate evidence for continuous improvement.

In practice, national cyber strategies increasingly converge around a shared logic that can be summarised as a resilience cycle: prevention and risk reduction; early detection and situational awareness; coordinated response and crisis management; and mitigation and recovery. The strategic emphasis is typically complemented by enabling pillars such as research and development, workforce development, public–private collaboration, and information sharing. This convergence is visible across major national strategies and regional initiatives, even where institutional models differ. As a result, cyber exercise programmes should be framed as a strategic implementation mechanism that validates whether these pillars work in real decision environments, especially at the interfaces between organizations and across sectors, where national-level resilience often fails in practice.

The U.S. National Cybersecurity Strategy (2023)[92] illustrates a strongly structured approach to strengthening national cyber defences. It emphasises investment in cybersecurity research and development, improved information sharing, and stronger cooperation between public and private actors as building blocks of a resilient cyber defence ecosystem. A prominent theme is the recognition that national preparedness requires a capable workforce; therefore, closing the cybersecurity skills gap and strengthening training capacity are treated as strategic priorities. The strategy also places explicit focus on the resilience of critical infrastructure and on protecting the

[92] The White House. (2023). National cybersecurity strategy. https://bidenwhitehouse.archives.gov/wp-content/uploads/2023/03/National-Cybersecurity-Strategy-2023.pdf

systems and networks that underpin essential services. Conceptually, this strategic framing can be expressed through a set of capability outcomes: resilience (withstanding and recovering from attacks), prevention, early detection, effective response, and impact mitigation. A recurring message in this policy logic is that success depends on multi-stakeholder engagement ,governmental, private sector, and civil society ,because national cyber resilience cannot be delivered by a single actor category.

The UK National Cyber Strategy (2022)[93] follows a comparable high-level structure, likewise emphasising resilience, prevention, detection, and response, and placing strong attention on research and development, information sharing, training, and resilience building. It also highlights the need to adapt to an increasingly sophisticated threat landscape, and it frames collaboration across government, private sector, and civil society as a central condition for national success. Additionally, international cooperation is presented as a key component: cyber threats cross borders, and effective response requires shared information, joint investigation capabilities, and coordinated action with partners. At the same time, strategy comparisons show that national documents differ in their degree of specificity regarding regulatory requirements and implementation instruments. Some strategies more clearly translate ambition into concrete compliance expectations (for example, around software supply-chain transparency requirements), while others remain at a more general policy level. This difference matters for national exercise governance because it affects what needs to be validated: where strategies include concrete obligations, exercise programmes can test compliance-ready execution (e.g., whether reporting, evidence handling, or supplier coordination works under time pressure); where strategies are less prescriptive, exercises can serve as a discovery tool to identify which implementation mechanisms are missing or under-developed.

At European level, recent initiatives increasingly add an operational and solidarity-oriented dimension to strategy. The EU Cyber Solidarity Act[94] conceptually combines a cyber emergency mechanism, a “cyber shield” based on collaborative Security Operations Centres, an incident review mechanism, and funding instruments. The cyber emergency mechanism emphasises preparedness through rigorous testing of entities in crucial sectors (such as finance, energy, and healthcare) in order to uncover gaps and vulnerabilities before real incidents occur. This is directly aligned with the core argument of this dissertation: exercises are not merely educational; they are a preparedness validation instrument. The EU Cybersecurity Reserve idea ,trusted service providers that can be deployed in major incidents ,also illustrates a strategic reliance on pre-arranged support capacity. Such mechanisms have immediate exercise implications: if a strategy assumes surge support and mutual assistance, exercises must test mobilisation, interface governance, access prerequisites, and the operational integration of external responders. Finally, the European Cyber Shield concept, built on cross-border collaboration between SOCs using advanced analytics and AI, highlights a strategy direction where detection and warning increasingly depend on shared sensing and rapid cross-border information exchange. An exercise programme aligned to such a strategy must therefore test not only technical detection, but also the governance of sharing: what can be shared, when, with whom, and how it is operationalized into coordinated action.

Across these strategy examples, several common components of a comprehensive cyber strategy can be distilled. First, strategies require a risk-driven orientation: systematic risk assessment and the use of threat intelligence to track evolving attacker behaviour and emerging vulnerabilities. Second, they establish policy frameworks that define priorities for protecting key assets and

[93] Cabinet Office. (2021). National cyber strategy 2022. HM Government. https://www.gov.uk/government/publications/national-cyber-strategy-2022

[94] European Parliament and Council of the European Union. (2025). *Regulation (EU) 2025/38 of 19 December 2024 laying down measures to strengthen solidarity and capacities in the Union to detect, prepare for and respond to cyber threats and incidents and amending Regulation (EU) 2021/694 (Cyber Solidarity Act). Official Journal of the European Union, L 2025/38*. EUR-Lex: Cyber Solidarity Act

essential services. Third, they emphasise collaboration, particularly between public and private sectors, because critical services and their supporting supply chains are often operated outside government. Fourth, they include incident response and cyber crisis management processes as core national capabilities, recognising that prevention and detection alone are insufficient. Fifth, they include capacity building through training and workforce development. Sixth, they rely on legal and regulatory frameworks that address cybercrime, data protection, and accountability, providing enforceable foundations for action. Seventh, they stress international cooperation and information exchange. Finally, they require continuous evaluation and improvement to remain robust against a changing threat landscape and technological advances, increasingly incorporating advanced technologies such as AI and analytics.

Hungarian military-scientific research has long recognised the relationship between cybersecurity, critical infrastructure protection, and national security. Kovács (2018) provides a comprehensive examination of cyber threats, cybersecurity arrangements, and the protection of critical information infrastructures, including their national and European dimensions.[95] Building on this foundation, the National Cyber Exercise Governance Model (C2) developed in this dissertation introduces a structured approach to translating strategic cybersecurity responsibilities into coordinated exercises, measurable preparedness indicators, and evidence-based capability improvement.

The evolution of national cybersecurity strategies is also demonstrated by Magyar et al. (2025)[96], who comparatively examined Hungary's cybersecurity strategies between 2013 and 2025. Their analysis identifies a transition from predominantly declarative strategic approaches towards more operationally oriented frameworks incorporating institutional development, societal participation, and alignment with EU and NATO cybersecurity expectations. These findings reinforce the relevance of the National Cyber Exercise Governance Model (C2), which translates national strategic objectives into coordinated exercise programmes, measurable capability assessment, and continuous resilience improvement.

For national cyber exercise programmes, the relevance of this strategic convergence is twofold. First, cyber exercises operationalise strategy by providing a controlled environment in which the effectiveness of strategic pillars can be tested. For example, a strategy that prioritizes public–private information sharing must be exercised through realistic sharing scenarios that include uncertainty, incomplete data, and legal constraints, because these are the conditions under which sharing mechanisms usually fail. A strategy that emphasises critical infrastructure resilience must be exercised in a way that stresses continuity trade-offs, dependency failures, and recovery prioritisation, not only technical containment. A strategy that emphasises workforce development must be linked to progressive exercise portfolios that build competence and validate skills over time, rather than treating exercises as isolated events.[97] Second, exercises provide structured feedback into strategy implementation: they reveal systemic friction points (unclear mandates, unrealistic assumptions, missing interfaces, or non-functional escalation paths) and generate evidence that can support prioritisation and investment decisions.

This dissertation therefore treats national cyber strategies as the strategic anchor for national cyber exercise governance. In the programme blueprint developed later in this chapter, strategic objectives are translated into cycle-level target objectives that define what should be exercised and

[95] Kovács, L. (2018). *A kibertér védelme*. Dialóg Campus Kiadó. https://www.uni-nke.hu/document/uni-nke-hu/Kov%C3%A1cs%20L%C3%A1szl%C3%B3.pdf

[96] Magyar, S., Bányász, P., Bányász-Váczi, K. B., & Pál, A. (2025). Magyarország kibervédelmi stratégiáinak összehasonlítása és fejlődési íve 2013 és 2025 között. *Szakmai Szemle, 23*(1), 127–144. https://ojs.knbsz.gov.hu/index.php/szakmai/article/view/228

[97] Radanliev, P. (2025). Cyber diplomacy: defining the opportunities for cybersecurity and risks from Artificial Intelligence, IoT, Blockchains, and Quantum Computing. *Journal of Cyber Security Technology*, *9*(1), 28-78.

measured in a given period. This enables a direct trace from national strategy to exercise objectives, and from exercise evidence to improvement actions. The approach also supports comparability across sectors and cycles: if exercises are aligned to stable strategic objective areas (e.g., resilience, detection, response, recovery, and collaboration), programme steering can evaluate coverage and progress systematically rather than relying on anecdotal reports. Finally, embedding exercises into national strategy strengthens legitimacy. It clarifies that exercises are not "optional training," but an operational assurance mechanism that helps a state demonstrate and improve cyber resilience across its critical functions.

In summary, national cyber strategies provide the authoritative framing for what a nation must be able to withstand, detect, coordinate, and recover from in cyberspace. Cyber exercises become vital when they are positioned as the practical mechanism to validate that strategic intent is executable under realistic constraints. When aligned to strategy, exercises can serve as repeatable system tests that generate auditable evidence and drive verified improvement, thereby linking policy ambition to measurable resilience outcomes.

Programme objectives and the national capability cycle

A national cyber exercise programme requires a stable objective hierarchy that bridges strategy and operational learning. At the highest level, the programme must be anchored in national cyber security and resilience objectives: ensuring continuity of essential services, strengthening crisis governance, improving cross-entity coordination, validating response and recovery capabilities, and increasing the ability to manage cascading consequences across sectors and supply chains. These strategic objectives should remain stable across years to enable longitudinal measurement. Their role is not to prescribe specific scenarios, but to define what "national resilience improvement" means and to provide a reference for prioritisation and evaluation.

National cyber capability should be understood as more than the availability of technical defensive tools. Kovács (2021) shows that cyber capabilities depend on the combined availability of organizational structures, trained personnel, operational processes, technological resources, and the ability to employ these elements coherently. For a national exercise programme, this implies that capability development must be assessed across organizational, human, procedural, and technical dimensions rather than through isolated technical performance indicators.[98]

Strategic objectives alone are insufficient for planning. They must be translated into cycle-level target objectives that are specific enough to be exercised and measured. Target objectives are set for a defined programme cycle (e.g., annually) and reflect current risk priorities, threat trends, systemic dependencies, and recurring gaps identified in previous cycles. The cycle-level layer is where improvement becomes enforceable: if exercises repeatedly identify the same coordination bottleneck, the next cycle's target objectives should explicitly include re-testing of that bottleneck under revised procedures, rather than simply selecting a new thematic scenario. In this sense, the national programme should be designed as a capability cycle, not as an event calendar. A capability cycle has four recurring stages: planning (direction setting and exercise portfolio definition), execution (conduct of exercises), evaluation (structured consolidation of findings), and improvement verification (tracking, remediation, and re-testing). The credibility of a national programme depends on the final stage; without verified improvement, exercises risk becoming performative compliance artefacts rather than instruments of resilience.

To ensure that programme objectives remain measurable rather than declarative, they should be translated into a structured evaluation logic. In this dissertation, this is supported by the ***Goal-Question-Metric (GQM)*** *approach and the* ***5-step Cyber Exercise Program (CXP) Evaluation Framework*** *developed in the author's own research.* This means that national objectives should

[98] Kovács, L. (2021). "Offenzív kiberműveletek II.: Kibererők és képességeik." Hadmérnök, 16(3), 119–137.

not only define what the programme intends to improve, but also which evaluative questions and indicators will later demonstrate whether improvement has actually occurred.

Governance model and institutional roles

A national exercise programme must define governance roles that enable coordination without over-centralisation. In general, three layers are required.

First, a **programme owner** (often a ministry, national cyber authority, or cross-government resilience body) provides strategic direction, mandates participation where applicable, and assures legitimacy through democratic accountability. The programme owner defines the minimum standard for programme exercises and maintains oversight of sensitive information handling.

Second, a **programme management office (PMO)** performs day-to-day coordination. Its responsibilities include maintaining the exercise portfolio and calendar, operating the programme platform, providing templates and guidance, supporting onboarding and maturity development, coordinating cross-sector exercises, and consolidating national-level evaluation outputs. The PMO should be small but professionally staffed; a programme that depends on ad-hoc volunteer facilitation will struggle to maintain continuity and comparability across cycles.

Third, **sector leads and exercise providers** translate national target objectives into sector-relevant exercises and ensure execution capacity. Sector leads may be regulators, sector coordination bodies, or designated critical infrastructure coordinators. Exercise providers can include national training centres, cyber range operators, academic or public–private competence centres, or certified private providers. This separation is important: the programme should not assume that the state must execute all exercises itself. Instead, it should set the rules of comparability and quality and allow multiple providers to deliver exercises under those rules.

This multi-actor governance logic is also consistent with Krasznay (2024)[99], who highlights the growing role of civilian cybersecurity companies in military cyber operations, particularly in relation to specialised expertise, technology, information sharing, and human resources. From a national cyber exercise perspective, this reinforces the need to integrate public, military, academic, and private-sector capabilities within a coordinated governance framework rather than treating national cyber capability as an exclusively governmental function.

Two additional governance mechanisms are necessary. A **steering board** (with representation from government, regulators, CERT/CSIRT communities, and critical sector stakeholders) ensures that target objectives remain nationally relevant and that the programme does not drift into narrow technical testing or purely bureaucratic reporting. A **quality and ethics function** (either embedded in governance or as an independent advisory role) ensures that exercises respect legal constraints, do not inadvertently create operational risk, and maintain proportionate data handling, particularly when exercises are connected to compliance, assurance, or national security considerations.

Exercise portfolio and progression model

A national programme must support a portfolio of exercise types rather than enforcing one “ideal” format. Different exercise formats test different capability layers: executive tabletops validate crisis governance and escalation decisions; functional exercises validate coordination processes and communication routines; technical simulations validate detection, containment, recovery and restoration under realistic operational constraints; and multi-organization exercises validate interdependency management and cross-entity collaboration. A portfolio approach allows the programme to address the full socio-technical nature of cyber crises.

[99] Krasznay, C. (2024). “The Role of Civilian Cybersecurity Companies in Military Cyber Operations.” Land Forces Academy Review, 29(1), 1–10.

The exercise portfolio should also remain threat-informed. Empirical analysis of contemporary conflicts demonstrates that cyber threat patterns, targets, and techniques evolve alongside geopolitical and operational developments (Bányász et al., 2024).[100] National programmes should therefore periodically review their scenario portfolio against observed threat trends so that repeated exercising does not become detached from the threat environment it is intended to prepare for.

To make a portfolio manageable at national scale, the programme should define a **progression model** that links exercise types to maturity and risk. At early maturity, organizations may begin with document reviews, communication drills, and small tabletops focused on roles, thresholds, and escalation. At intermediate maturity, functional exercises and cross-department staff exercises validate procedures and coordination under time pressure. At higher maturity, range-enabled technical exercises validate operational response and recovery capabilities, ideally including degraded modes, supply-chain complications, and service continuity requirements. The progression model enables national planners to assign exercise expectations proportionately: not every participant needs to join a full-scale national technical simulation every year, but every participant should demonstrate baseline readiness and continuous improvement through an appropriate exercise path.

A programme also benefits from explicitly distinguishing three exercise categories:

1. **Organization-level exercises** (single entity): build internal capability and validate governance and operational readiness.
2. **Sector-level exercises** (multi-entity within one sector): validate sector coordination, shared dependencies, and regulator interfaces.
3. **National-level exercises** (cross-sector): test whole-of-society coordination, national situational awareness, inter-sector dependencies, and public communication dynamics.

By treating these categories as complementary, the programme can build depth (organization maturity) and breadth (system coordination) in parallel. Importantly, national-level exercises should not be the only “flagship” activity. If national exercises are run infrequently and without systematic follow-up, their learning effects are limited. A sustainable programme therefore uses national exercises as capstone tests of the ecosystem while embedding continuous learning primarily through organization and sector exercises.

Standardized artefacts and a minimum dataset for comparability

Comparability and auditability require a minimum standard for exercise artefacts. Standardization does not mean that all exercises must look identical. It means that each programme exercise must produce a defined set of deliverables and metadata such that an external reviewer ,whether a regulator, an auditor, or the programme PMO ,can understand what was planned, what was tested, what evidence was collected, and what improvements were assigned.

A pragmatic minimum artefact set for programme exercises includes:

- **Exercise concept / exercise plan**: purpose, scope, assumptions, boundaries, roles, target audience, logistics, and constraints.
- **Scenario backbone**: narrative context, threat model, progression phases, key decision points, and links to objectives.
- **Inject plan / MSEL** (where applicable): timeline of events and injects, mapping injects to objectives, target roles, expected actions, and observation points.

[100] Bányász, P., Bottyán, S., Dobó, J., Dub, M., Gyaraki, R., et al. (2023). “Biztonságorientáció és -tudatosítás nagyvállalati környezetben.” In Nagyvállalati biztonságtudatosság és nyílt forrású információszerzés (pp. 8–80). Ludovika.

- **Evaluation plan**: what will be measured, by whom, using which methods, and how evidence will be collected and protected.
- **After-action report**: consolidated findings, lessons identified, and an improvement plan with owners, deadlines, and verification approach.

The minimum dataset should support not only description and comparability but also structured evaluation. For this reason, the dataset should contain fields that enable Goal-Question-Metric-based analysis, including programme objective linkage, exercise questions or evaluation focus, chosen indicators, data collection methods, and reporting outputs. In this way, standardized artefacts do not merely document exercises; they also enable defensible programme measurement.

For national comparability, these artefacts must be accompanied by a **minimum dataset**: structured metadata that can be aggregated without exposing sensitive operational details. Typical metadata elements include exercise type and level (discussion-based, functional, technical; organizational/sector/national), sectors involved, number of participating organizations, duration and mode (onsite/distributed/hybrid), objectives and objective categories, scenario theme and threat class (e.g., ransomware, supply-chain compromise, OT disruption), and evaluation method category. The minimum dataset is not a replacement for rich narratives; it is an indexing layer that allows national stakeholders to answer questions such as: Which sectors have exercised cross-entity coordination this year? Which objective areas are undertested? Which systemic gaps recur across sectors? Which improvement themes show progress over time?

Confidentiality and trust are decisive for participation. The programme should therefore adopt a **tiered reporting model**. A full internal after-action report may include sensitive details and performance data; a sector-level summary may anonymize participants; a national-level summary may further abstract findings into themes and metrics. This tiering allows the programme to balance learning and accountability with realistic constraints on what organizations will share. It also supports public communication: in many contexts, demonstrating that a country invests in resilience is valuable, but publishing tactical weaknesses or sensitive dependencies is not acceptable.

The national cyber exercise platform as enabling infrastructure

A national exercise programme requires a platform that functions as institutional memory, coordination backbone, and measurement layer. Without a platform, exercises remain scattered across organizations, and national learning depends on informal relationships or one-off reporting. The platform should not be conceived as a complex “cyber range system,” but as an information and workflow infrastructure with clear governance.

A robust platform typically provides five capability clusters:

1. **Exercise registry and calendar**: a national overview of planned and completed exercises, filtered by sector, level, and type; includes a mechanism for coordinating dates to avoid overload and to enable cross-sector participation.
2. **Artefact repository and templates**: standardized templates, guidance documents, checklists, and example artefacts; supports versioning to ensure traceability and to enable reuse across cycles.
3. **Submission and reporting workflows**: structured upload or form-based submission of the minimum dataset and required artefacts; supports tiered confidentiality, access control, and audit logs.

4. **Evaluation and analytics dashboard**: aggregation of metadata and high-level findings; trend analysis across cycles; ability to track objective coverage, participation patterns, recurring gap themes, and improvement status.
5. **Improvement tracking and verification**: a mechanism to register action items, assign owners, set deadlines, capture evidence of completion, and schedule follow-up verification exercises or tests.

The platform should also support the operationalization of the CXP Evaluation Framework. In practice, this means enabling structured capture of programme goals, evaluation questions, selected metrics, lifecycle-generated data, and improvement actions. A platform that stores only reports is insufficient; it should function as a measurement backbone that supports data collection and analysis across programme management, exercise design and development, exercise conduct, and exercise evaluation.

The platform's value is not primarily technological; it is organizational. It lowers the barrier for participation by providing ready-to-use structure, reduces heterogeneity in artefact quality, and makes national-level analysis possible without requiring manual consolidation. If the programme includes an improvement tracking function, the platform becomes the anchor for the most important part of the national capability cycle: ensuring that lessons identified are converted into measurable change.

From a security and trust perspective, the platform must implement strict access control, robust audit logging, and clear information classification and sharing rules. Even when exercises are not classified in a national security sense, they often contain sensitive operational information, details of dependencies, and performance observations that could be misused or politically misinterpreted. A platform that cannot credibly protect this information will undermine participation and may produce an incentive to submit minimal and low-quality data. Therefore, platform governance should explicitly define who can access what data, for which purpose, and under which confidentiality agreements.

Measurement: From exercise completion through KPIs to resilience evidence

A recurring weakness of cyber exercise programmes is over-reliance on completion metrics such as the number of exercises conducted, the number of participants involved, or general satisfaction feedback after the event. These indicators are useful for programme administration and visibility, but they do not demonstrate whether the programme actually improves cyber resilience at national level. A national cyber exercise programme therefore requires a measurement framework that moves beyond exercise completion and links programme activity to capability development, learning quality, and verified improvement over time. This is especially important in defence-oriented and critical-infrastructure settings, where exercises are increasingly expected to serve not only as training events, but also as evidence of preparedness, coordination capability, and resilience governance.

The challenge is that national Cyber Exercise Programs (CXPs) are inherently difficult to evaluate in a consistent manner. Different exercise cycles often involve different participants, different sectors, different threats, and different capability targets. Large-scale national exercises may change focus from year to year, for example from healthcare to energy, from crisis communication to third-party dependencies, or from strategic coordination to operational recovery. As a result, direct comparison between exercises is often methodologically weak if no common evaluation logic exists. Previous research by Páhi, Tóth and Tjoa (2025)[101] identified the absence of standardized evaluation approaches as a central limitation of national cyber exercise programmes and proposed

[101] Páhi, T., Tóth, A., & Tjoa, S. (2025). Evaluating national cyber exercise programs. In Proceedings of the 19th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI 2025). IEEE.

a five-step evaluation framework based on the Goal-Question-Metric paradigm. The present dissertation builds upon and extends this work by embedding the evaluation logic within a broader national cyber exercise governance and capability-development model. The paper's central contribution is the development of a structured **5-step CXP Evaluation Framework** based on the **Goal-Question-Metric (GQM)** paradigm, specifically designed for the evaluation of national cyber exercise programmes.

The relevance of this contribution for the present dissertation is direct. The dissertation argues that a national cyber exercise programme must function as a measurable capability cycle rather than as a loose collection of events. The paper provides exactly the methodological bridge needed for this argument: it translates the abstract requirement of "measurable resilience improvement" into a structured evaluation process. In this sense, the paper strengthens the dissertation's governance model by adding a more formal evaluation logic that can be embedded into programme governance, standardized artefacts, and after-action improvement loops.

The proposed measurement framework therefore combines two complementary perspectives. First, it retains the dissertation's three-layer logic of measurement at programme level, exercise level, and improvement-cycle level. Second, it strengthens this logic through the GQM-based methodology of the paper, which provides a reproducible way to derive indicators from programme goals rather than selecting metrics in an ad hoc manner. This is important because national cyber exercise measurement should not start with "available data," but with clearly defined national and programme objectives.

A Goal-Question-Metric logic for national cyber exercise programmes

The Goal-Question-Metric approach provides a structured measurement model consisting of three levels. At the first level, a goal is defined for the object of evaluation. At the second level, questions are formulated that operationalize what it means to achieve that goal. At the third level, quantitative or structured qualitative metrics are selected to answer these questions in a measurable way. In the context of national cyber exercise programmes, this means that evaluation begins with strategic and programme-level intent rather than with isolated exercise observations. In other words, the programme first defines what it seeks to improve, for example cross-sector coordination, incident escalation, continuity decision-making, or recovery governance; it then formulates questions that make these ambitions evaluable; and only after that does it define the concrete metrics and data collection methods. This makes the evaluation logic more transparent, more auditable, and more suitable for longitudinal comparison across cycles.

*The **5-step CXP Evaluation Framework**, consisting of:*
(1) defining goals,
(2) formulating questions,
(3) identifying metrics,
(4) conducting data collection and analysis, and
(5) reporting findings and feeding them into improvement.

For the present dissertation, this 5-step model is highly valuable because it aligns naturally with the governance logic of a national programme. Goal definition corresponds to national strategy and programme objectives; question formulation corresponds to exercise design and evaluation planning; metrics identification connects to the minimum dataset and standardized artefacts; data collection and analysis correspond to exercise execution and evaluation; and reporting and improvement correspond directly to the dissertation's after-action and remediation loop.

Three measurement layers for resilience-oriented programme governance

Within the dissertation, a pragmatic national measurement model distinguishes three interrelated layers.

First, programme-level metrics capture coverage, engagement, structure, and strategic alignment across the national exercise portfolio. These indicators do not yet prove resilience outcomes, but they show whether the programme is reaching the intended ecosystem and whether it is evolving in the desired direction. Examples include the number and growth of exercises over time, the number of exercises per year, the distribution of exercise types, the frequency of cross-entity or cooperation-focused exercises, participation by sector, maturity distribution, availability of subject-matter experts, use of standard templates or programme tools, and the degree to which exercises are aligned with national cybersecurity priorities. The paper highlights several such programme-management-oriented metrics and thereby strengthens the dissertation's claim that national governance requires visibility not only into single exercises, but also into the portfolio as a whole.

Second, exercise-level performance indicators assess how well a specific exercise achieved its stated objectives. Here, metrics should be objective-linked and context-sensitive. They may include response time to escalation, time to identify systems at risk, decision timeliness, completeness of situational reporting, communication effectiveness, restoration planning quality, number of played versus planned injects, channels used for inject delivery, duration of real and virtual exercise components, number of facilitators and observers, or the percentage of exercise goals achieved. The paper also highlights scenario credibility, perceived participant stress, feedback scores, and the number of actionable recommendations in the evaluation report as potentially useful indicators. These metrics are especially relevant because they connect exercise design and conduct to observable performance rather than relying solely on narrative judgement. At the same time, the dissertation's caution remains valid: such metrics must not be used as simplistic league tables. In national programmes, they should serve learning, comparison of patterns, and improvement tracking rather than punitive ranking.

Third, improvement metrics across cycles provide the closest proxy for resilience enhancement, because they focus on whether identified findings are actually translated into corrective action and later verified. This includes indicators such as the percentage of action items completed on time, the proportion of corrective actions re-tested through follow-up exercises, reduction in recurrence of systemic weaknesses, increased objective coverage in previously under-tested domains, and the degree to which previously observed capability targets are later reached with fewer or smaller challenges. This layer is particularly important for the dissertation because the central argument is that national resilience does not improve simply because exercises are conducted. It improves when repeated cycles of exercising produce evidence of capability change.

Using lifecycle-generated data as measurement input

A particularly useful contribution of the paper is that it does not treat evaluation as something that starts only after the exercise. Instead, it emphasises that data relevant for indicators and metrics are generated throughout the cyber exercise lifecycle. This is a critical insight for the dissertation, because it aligns measurement with lifecycle governance rather than with post hoc reporting only. The paper structures metric-relevant data along exercise lifecycle phases derived from HSEEP and highlights that meaningful indicators can be collected from programme management, exercise design and development, exercise conduct, and exercise evaluation.

This lifecycle-oriented measurement perspective should be adopted in the national programme blueprint. From the **programme management** phase, data can be collected on exercise growth, budget, exercise typology, yearly distribution, and supporting expert capacity. From **exercise design and development**, data can be captured regarding scope, participant composition, sectors involved, exercise series, scenario structure, inject planning, planned duration, use of tools, and alignment between objectives and programme priorities. From **exercise conduct**, the programme can record how the scenario unfolded in practice, including deviations, inject delivery, participant response timing, facilitation structure, and operational tempo. Finally, from **exercise evaluation**,

the programme can capture achieved goals, capability ratings, stress and realism indicators, data collection methods used, and the number and quality of recommendations generated. This is methodologically powerful because it enables the programme to trace how design choices influence conduct and how conduct influences findings, rather than evaluating outcomes in isolation.

Possible KPI groups for a national cyber exercise programme include the following:

(1) **programme management KPIs**, such as exercises per year, growth over time, sectoral coverage, cross-sector exercise rate, distribution by exercise type, participation breadth, budget, and use of standard tools;
(2) **exercise design KPIs**, such as alignment of objectives with programme priorities, duration of planning, number of planned participants, use of exercise-management tools, number of injects, delivery modality, and whether the exercise is part of a recurring series;
(3) **exercise conduct KPIs**, such as played injects versus planned injects, number of facilitators and observers, average participant response time, time to identify affected systems, time to escalate, scenario adherence rate, inject channels used, and total exercise runtime;
(4) **exercise evaluation KPIs**, such as percentage of achieved exercise goals, capability targets reached, data collection methods used, perceived participant stress, scenario credibility, feedback scores, and number of actionable recommendations; and
(5) **improvement-cycle KPIs**, such as number of corrective actions created, corrective action completion rate, time to close findings, re-test rate, recurrence rate of systemic weaknesses, and longitudinal improvement in capability, coordination, detection, escalation, and recovery performance.

To translate the above measurement logic into an operational programme instrument, *Table 11* proposes a structured set of possible national cyber exercise KPIs. The table builds on the metrics identified in the author's paper on Cyber Exercise Program evaluation and extends them for dissertation use by organizing them into programme-level, design-level, conduct-level, evaluation-level, and improvement-cycle indicators.

KPI category	KPI	What it measures	Example unit	Why it matters
Programme management	Number of exercises per year	Overall programme activity level	Absolute number per year	Shows whether the programme is sustained over time
	Growth of exercises over time	Expansion or contraction of the programme portfolio	% increase/decrease year-on-year	Indicates programme maturity and scaling
	Sectoral coverage	Extent to which critical sectors are included	Number / % of priority sectors covered	Supports national representativeness
	Cross-sector exercise rate	Share of exercises involving more than one sector or entity type	% of exercises	Measures ecosystem-level coordination focus
	Distribution by exercise type	Balance between tabletop, functional, technical, and integrated exercises	% by type	Shows whether the portfolio supports progression and maturity

	Repetition / cadence of exercises	Whether exercising is periodic rather than ad hoc	Annual / semi-annual / one-off	Indicates institutionalisation
	Number of participating organizations	Breadth of institutional engagement	Absolute number	Reflects national reach
	Availability of subject-matter experts	Expert support capacity for planning and evaluation	Number of SMEs / evaluator pool size	Indicates programme sustainability
	Budget or funding per cycle	Resources available for national exercise delivery	Currency / annual budget	Supports feasibility and scale assessment
	Use of standard templates and tools	Degree of methodological standardisation	Yes/no or % of exercises using them	Improves comparability and auditability
	Public visibility / communication output	External signalling and awareness dimension	Press release yes/no; number of public outputs	Relevant for transparency and national awareness
Exercise design and development	Objectives aligned with programme priorities	Strategic alignment of each exercise	% of exercises with explicit linkage	Prevents disconnected event planning
	Objectives aligned with national strategy	Whether exercise design reflects national resilience priorities	Yes/no; % aligned	Connects exercises to policy goals
	Duration of planning	Planning effort and maturity	Days / weeks / person-days	Useful for efficiency comparison
	Use of exercise-management tools	Degree of methodological and technical support in preparation	Yes/no; tool type	Supports repeatability and structured design
	Number of participants planned	Planned scale of participation	Absolute number	Indicates complexity and scope
	Participant diversity	Breadth of roles, organizations, and sectors included	Number of actor categories represented	Supports realism and coordination testing
	Number of planned injects	Planned scenario intensity and control granularity	Absolute number	Indicates scenario sophistication
	Planned dynamic inject rate	Share of injects that may be adapted during conduct	% of injects	Relevant for realism and adaptive control
	Exercise duration (planned)	Planned operational tempo	Hours / days	Helps compare exercise burden
	Real vs. virtual delivery model	Modality of the exercise	Real / virtual / hybrid	Supports typology-based comparison
	Series vs. one-off exercise	Position of the event within a longitudinal programme	Binary / frequency	Important for follow-up and improvement logic
Exercise cond	Played injects versus planned injects	Extent to which the scenario was executed as intended	% delivered	Indicates execution fidelity

	Number of facilitators / observers / counterplayers	Delivery and evaluation capacity during execution	Absolute number	Important for exercise quality and evidence capture
	Average participant response time	Speed of operational reaction to scenario events	Minutes / hours	One of the strongest operational indicators
	Time to identify affected systems	Detection and situational-awareness speed	Minutes / hours	Strong resilience-relevant signal
	Time to escalate	Speed of internal or cross-entity escalation	Minutes / hours	Critical for crisis coordination
	Scenario adherence rate	How much of the intended scenario unfolded as planned	% without major deviation	Indicates execution control and design quality
	Channels used for inject delivery	Delivery realism and communication complexity	Email / phone / platform / media / mixed	Supports scenario and method analysis
	Methods of inject delivery	Variety of exercise stimulation mechanisms	Manual / scripted / technical / media / hybrid	Useful for typology comparison
	Total exercise runtime	Actual execution burden	Hours / days	Supports efficiency and comparability
	Cross-entity coordination events observed	Number of observable inter-organizational decision points	Absolute number	Supports national coordination analysis
Exercise evaluation	Percentage of achieved exercise goals	Goal fulfilment at exercise level	% achieved	Core KPI directly linked to objectives
	Capability targets reached	Extent to which intended capabilities were demonstrated	Number / % reached	Supports capability-based assessment
	Performance rating by objective	Quality of performance per objective	Scale, e.g. achieved / partly achieved / major challenges	Supports structured comparison
	Data collection methods used	Evaluation robustness and triangulation	Observation / survey / interview / hotwash / doc review	Indicates methodological quality
	Participant stress perception	Realism and pressure level	Likert scale / yes-no	Useful proxy for immersion
	Scenario credibility	Perceived realism and plausibility	Likert scale	Important for training validity
	Feedback score per question	Participant or evaluator assessment of exercise quality	Average score	Supports structured post-exercise comparison

	Number of actionable recommendations	Practical value of the exercise output	Absolute number	Strong bridge to improvement planning
	Media or external reporting triggered	Visibility and communication relevance	Yes/no	Useful for strategic or public-sector exercises
Improvement cycle	Corrective actions created	Translation of findings into action	Absolute number	Shows whether findings were operationalized
	Corrective action completion rate	Share of actions implemented on time	% completed	Core resilience-improvement KPI
	Time to close corrective actions	Speed of organizational learning	Days / weeks / months	Indicates follow-through discipline
	Re-test rate of previous findings	Whether remediation is actually verified later	% of actions re-tested	Strong indicator of maturity
	Recurrence rate of systemic weaknesses	Whether the same problems reappear across cycles	% recurring findings	Central national learning indicator
	Improvement in capability score over time	Longitudinal capability development	Delta between cycles	Strongest proxy for resilience growth
	Improvement in coordination performance over time	Progress in multi-actor response quality	Delta / trend	National ecosystem relevance
	Reduction in time-to-detect / escalate / recover	Verified operational improvement	Trend over repeated exercises	Demonstrates measurable resilience gain

TABLE 12 - POSSIBLE KPIS FOR A NATIONAL CYBER EXERCISE PROGRAMME

The KPI set above should not be interpreted as a rigid universal scorecard. Rather, it provides a structured menu of indicators from which a national programme can derive a context-appropriate KPI architecture based on programme goals, exercise type, sectoral sensitivity, and data availability. In line with the Goal-Question-Metric logic, not all exercises must use all indicators. However, a stable national core set should be maintained to enable longitudinal comparison, confidentiality-aware aggregation, and evidence-based resilience improvement across exercise cycles. The indicators are derived from and extended beyond the metrics proposed in the author's Cyber Exercise Program evaluation paper, which highlights lifecycle-based measurement across program management, exercise design and development, exercise conduct, and exercise evaluation.

From national KPIs to evidence-informed interpretation

For national comparability, the selection of indicators must prioritize robustness, clarity, and interpretability over excessive statistical complexity. Not every valuable aspect of an exercise can be measured numerically, and not all organizations operate under the same legal, technical, or operational constraints. This means that national KPI design should follow two principles.

The first principle is **objective linkage**: every metric should be traceable to a defined programme or exercise goal through an explicit question. This is where the GQM method is particularly valuable, because it forces the evaluator to explain why a metric is being collected and what it is intended to show. The second principle is **context-aware comparability**: metrics should be sufficiently standardised to support aggregation and trend analysis, but interpreted carefully in light of organizational maturity, sectoral constraints, and exercise type. A national programme can

therefore compare patterns and trajectories without pretending that all exercises are directly interchangeable.

This is especially important in defence-oriented and mixed civilian-military settings. A strategic crisis-management exercise involving ministries, a sectoral continuity exercise involving critical operators, and a technical range-based defence exercise will not generate identical kinds of evidence. However, they can still be connected through a structured national KPI architecture if the programme records common metadata, maps exercise objectives to shared resilience dimensions, and uses tiered measurement logic. In this sense, the dissertation's KPI framework and the paper's GQM-based evaluation method are mutually reinforcing: the KPI framework defines what should matter at national level, while the GQM process defines how those indicators should be derived, interpreted, and improved.

The integration of the paper into this chapter also has methodological significance for the dissertation as a whole. The dissertation's hypotheses on national governance and measurable resilience require more than conceptual argument; they require a plausible evaluation mechanism through which national improvement can actually be assessed. The paper contributes such a mechanism. It offers a structured and reusable evaluation methodology that can support longitudinal analysis of national cyber exercise programmes and thereby strengthens the evidentiary basis of the dissertation's governance proposal. It also contributes an initial catalogue of concrete metrics that can later be refined into the dissertation's national KPI set.

Moreover, the paper supports the dissertation's broader claim that standardisation is not only a governance benefit, but also a scientific necessity. If programme goals, evaluation questions, metrics, and data collection approaches are not standardised to a minimum degree, neither national learning nor research evaluation is possible in a robust way. The GQM-based CXP Evaluation Framework therefore serves as both a practical governance tool and a research-oriented measurement model.

Conclusion

In sum, the measurement framework of a national cyber exercise programme must evolve from simple completion reporting toward resilience-oriented evidence generation. The author's paper *Evaluating National Cyber Exercise Programs* makes an important contribution in this respect by introducing a structured 5-step CXP Evaluation Framework based on the Goal-Question-Metric approach and by identifying lifecycle-linked metric categories suitable for national programme evaluation. Integrated into the dissertation's programme blueprint, this contribution strengthens the transition from exercises as isolated events to exercises as measurable elements of a national capability cycle. It provides the methodological foundation for deriving national KPIs from programme goals, structuring data collection across the lifecycle, and linking exercise observations to longitudinal evidence of preparedness, capability growth, and verified resilience improvement. In this way, measurement becomes not a reporting afterthought, but a central governance mechanism through which national cyber exercise programmes can demonstrate value, justify investment, and support continuous resilience enhancement over time.[102]

Institutionalising after-action improvement loops

The decisive differentiator between an exercise programme and a set of exercises is the improvement loop. After-action reviews are common; sustained remediation and verification are not. A national programme should therefore mandate an **Improvement Plan** as a structured output of each programme exercise, with clearly assigned ownership, deadlines, prioritization logic, and

[102] Pahi, T., Toth, A., & Tjoa, S. (2025, May). Evaluating National Cyber Exercise Programs. In 2025 IEEE 19th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 000553-000558). IEEE.

evidence expectations. The programme PMO should provide an action plan template and an evidence catalogue model that makes follow-up manageable.

Improvement governance requires clarity on two categories of findings. First, there are **capability findings**: gaps in response processes, coordination, tooling, staffing, decision-making, or inter-organizational agreements. Second, there are **exercise design findings**: issues with scenario plausibility, inject delivery, evaluation design, participant briefing, or data collection. Both categories matter, but they have different owners and different remediation paths. If a programme mixes them without distinction, improvement tracking becomes noisy and less credible.

Verification is the final step and should be operationalized through proportionate re-testing. Not every finding requires a national-level replay. Many improvements can be validated through targeted mini-drills, focused tabletops, technical tests in a lab environment, or specific re-runs of a decision point. The programme should encourage a “small but frequent verification” culture: low-overhead follow-ups that prevent the improvement loop from being deferred to the next annual cycle. Over time, this builds institutional discipline and increases confidence that exercise outputs lead to actual change.

Within the 5-step CXP Evaluation Framework, reporting and improvement form the final step and ensure that measurement results are translated into corrective action, tracked remediation, and follow-up verification.

Implementation roadmap: building the programme in phases

Because national programmes operate under resource constraints and diverse maturity, implementation should be phased.

Phase 1: Establish the minimum standard and governance. Define programme scope, strategic objectives, participation model, confidentiality rules, minimum artefact set, and minimum dataset. Set up the PMO and a steering mechanism. Identify pilot sectors and a limited number of programme exercises.

Phase 2: Launch a pilot portfolio and the platform foundation. Run a small set of programme exercises across selected sectors with strict artefact and reporting compliance. Implement the platform’s core functions: registry/calendar, templates, and reporting workflows. Use the pilot to calibrate templates, evaluator methods, and confidentiality tiers.

Phase 3: Expand participation and introduce improvement tracking. Broaden the programme to additional sectors and organization types. Add improvement plan tracking to the platform and begin cross-cycle analysis. Introduce evaluator training and a basic measurement model.

Phase 4: Mature into a capability cycle with verified improvement. Institutionalise follow-up verification and integrate the programme into broader resilience governance and assurance practices. Run cross-sector capstone exercises periodically, using accumulated data to select nationally meaningful objectives and to re-test systemic dependencies.

This phased approach acknowledges that programme success depends on trust, usability, and demonstrable value. A national programme that demands heavy reporting before providing enabling support will face compliance fatigue. Conversely, a programme that provides support without enforcing minimum standards will not achieve comparability and national learning.

Scientific Contribution and Added Value of the National CX Governance Model

The principal contribution of this chapter is the transformation of the national cyber exercise programme from a general policy aspiration into a structured and implementable governance model. The analysis developed throughout Section 2.4 demonstrates that national cyber resilience cannot be strengthened in a sufficiently systematic or evidential manner through isolated exercise

activities alone. Individual exercises can generate valuable observations, identify weaknesses, and support organizational learning, but their contribution to national resilience remains limited if they are not embedded within a programme architecture that connects strategic priorities, governance responsibilities, exercise design, comparable information structures, evaluation, corrective action, and subsequent verification across repeated exercise cycles. The National Cyber Exercise Governance Model therefore conceptualises cyber exercising as a sustained national capability-development process rather than as a sequence of discrete events.

A central aspect of the contribution is the introduction of a **programme perspective** that links national strategic intent to operational exercise activity. Within the proposed model, strategic cybersecurity and resilience priorities are translated into programme-level objectives, which subsequently inform the composition of the national exercise portfolio and the objectives of individual exercise activities. Exercise outputs are then evaluated against these objectives and converted into findings, corrective actions, and re-testing requirements. This creates the basis for longitudinal learning and enables national stakeholders to move beyond activity-oriented questions, such as whether an exercise was conducted, towards capability-oriented questions concerning whether known weaknesses have been reduced, whether systemic gaps recur across sectors, and whether observable evidence indicates improvement over time.

The governance model further contributes by defining the institutional functions required to sustain such a programme. National cyber exercise governance necessarily involves actors with different mandates, levels of authority, sectoral responsibilities, and information-access conditions. The proposed model therefore follows a federated rather than purely centralized logic. It distinguishes between strategic programme ownership, programme management and coordination, sectoral or domain-level responsibility, individual exercise ownership and provision, participating organizations, and evaluation or assurance functions. This division of responsibilities is intended to provide sufficient central coherence for national learning while preserving the operational autonomy of sectoral and organizational actors. The model is therefore particularly suited to national cybersecurity environments in which responsibility is institutionally distributed and where excessive centralization would be neither realistic nor desirable.

A further scientific contribution concerns **comparability**. The chapter operationalizes the Cyber Exercise Taxonomy and Minimum Dataset developed in Section 2.3 by embedding them within programme governance. The purpose of these mechanisms is not to standardise exercise scenarios or execution formats, but to standardize the information required to understand and compare heterogeneous exercise activities. This distinction is essential. Strategic tabletop exercises, technical live-fire exercises, sectoral functional exercises, and multinational crisis-management exercises should remain different where their purposes differ. Nevertheless, they can still contribute to one national programme if they are described through common metadata, linked to defined capability objectives, documented through compatible artefact structures, and evaluated through sufficiently consistent evidence logic.

The resulting information architecture creates the basis for portfolio-level visibility. National stakeholders can identify which sectors, capability areas, exercise types, and maturity levels have been covered, where exercise activity is concentrated, which gaps remain untested, and whether similar findings recur across organizations or programme cycles. At the same time, the model explicitly recognises that national comparability must not require unrestricted centralization of operationally sensitive information. Programme-level metadata, selected controlled artefacts, and restricted operational evidence can be governed through differentiated access arrangements. In this way, the model seeks to reconcile comparability with confidentiality, classification, data minimization, and the need-to-know principle.

The contribution is further strengthened by the formal integration of **measurement and evaluation** into the governance model. The dissertation incorporates the author's earlier research on national Cyber Exercise Programme evaluation and the five-step CXP Evaluation Framework based on the

Goal-Question-Metric approach. This framework links programme objectives to evaluative questions and subsequently to measurable indicators, thereby creating a structured methodology for national KPI design, lifecycle-based data collection, and longitudinal assessment. The importance of this integration lies in the distinction between exercise activity and resilience evidence. Indicators such as the number of exercises, participating organizations, or exercise hours may describe programme activity, but they do not in themselves demonstrate capability improvement. The proposed model therefore places greater emphasis on objective-linked performance indicators, recurring findings, corrective-action closure, time to remediation, re-test results, and other evidence that can demonstrate whether weaknesses identified through exercising have been addressed[103].

The most consequential element of the contribution is the institutionalization of the **after-action improvement loop**. The model treats an After-Action Report or lessons-identified process as an intermediate rather than final output of exercising. A finding becomes relevant to resilience development only when it is assigned to a responsible actor, translated into a corrective action, implemented, monitored, and subsequently verified. Re-testing therefore becomes a constitutive element of the programme rather than an optional follow-up activity. The model allows such verification to be proportionate to the significance and nature of the finding; it may take the form of a focused tabletop exercise, a technical test, a targeted drill, or a later full-scale exercise. What matters is that the programme preserves the logical connection between evidence, finding, remediation, and verification.

This improvement logic is the main mechanism through which the proposed governance model addresses the second research gap identified in the dissertation. The model does not assume that exercising automatically produces resilience. Instead, it proposes a governance architecture through which exercise-generated evidence can be translated into verified capability development. The relevant causal claim is therefore deliberately bounded. The dissertation does not argue that the existence of a national exercise programme is sufficient to increase national cyber resilience in all contexts. Rather, it argues that a programme built around defined governance, comparable information, objective-linked measurement, corrective-action ownership, and re-testing provides a more suitable and auditable framework for resilience-oriented improvement than fragmented or ad hoc exercise activity.

The scientific added value of the National Cyber Exercise Governance Model therefore lies primarily in **integration**. Existing national and international practice already contains many of the individual elements used in the model, including national strategies, exercise programmes, lifecycle methodologies, after-action processes, metrics, and sectoral coordination arrangements. The contribution of the dissertation is not to claim that these components are individually new. Its originality lies in combining them into a coherent national architecture in which strategy, portfolio governance, taxonomy, common artefacts, minimum datasets, evaluation, information governance, corrective action, and re-testing are treated as interdependent components of one capability-development system.

In this sense, the model constitutes the principal constructive response to H2 and MRQ2. It provides a governance-ready and adoption-oriented framework through which heterogeneous national cyber exercise activities can be made more comparable, auditable, evidence-oriented, and improvement-focused without requiring uniform exercise design or excessive centralization. The contribution can therefore be understood as a shift from **exercise management** to **national exercise governance**, and from the completion of individual exercises to the systematic management of capability development across exercise cycles

[103] Páhi, T., Tóth, A., & Tjoa, S. (2025, May). *Evaluating National Cyber Exercise Programs.* In *2025 IEEE 19th International Symposium on Applied Computational Intelligence and Informatics (SACI)* (pp. 000553–000558). IEEE.

Position of the Programme Concept within the Dissertation Architecture

Within the broader architecture of the dissertation, the national cyber exercise programme serves a function that extends beyond the practical organization of national exercise activity. It acts as the principal **methodological and governance scaffold** connecting the analytical findings of Part I with the lifecycle and AI-operationalization contributions that follow. The programme concept therefore occupies an intermediate position in the cumulative research logic: it translates the external baseline identified at international and European level into a national governance structure within which individual exercises, exercise data, evaluation mechanisms, and later AI-supported processes can operate coherently.

The programme model is built on the results of the preceding analytical stages. Part I establishes that international and European cyber exercise practice remains heterogeneous, while convergence is emerging around governance, lifecycle orientation, structured artefacts, evidence, evaluation, dependency awareness, and continuous improvement. Section 2.2 then examines Austria as the principal national reference case and demonstrates how European requirements encounter a distributed national environment composed of ministries, competent authorities, security and defence actors, CSIRTs, regulators, critical entities, and private-sector stakeholders. Section 2.3 subsequently converts the comparability problem into a formal Cyber Exercise Taxonomy and Minimum Dataset. The programme concept developed in Section 2.4 integrates these analytical outputs and translates them into a national capability architecture.

For this reason, the programme should not be interpreted as an additional exercise methodology. Its role is different. It defines the **governance environment** in which exercise methodologies are applied. It determines how strategic priorities become programme objectives, how exercises are selected and positioned within a portfolio, how actors are coordinated, which minimum information and artefacts are required, how evidence is evaluated, and how findings are converted into corrective action and re-testing. The programme concept therefore establishes the institutional and informational conditions necessary for repeatability, comparability, and national learning.

This distinction becomes particularly important in relation to Section 2.5. The ENISA-based AI-Enhanced Cyber Exercise Lifecycle is not intended to replace the programme model, nor does the programme model prescribe the detailed process through which each exercise is designed and conducted. The two contributions operate at different but connected levels. The programme defines **what must be governed and why**, while the lifecycle defines **how an individual exercise progresses through a repeatable process**. The programme provides the national objectives, governance requirements, artefact expectations, information constraints, and evidence logic; the ENISA-based lifecycle operationalizes these requirements through the phases of Initiation, Design, Preparation, Execution, Evaluation, and Moving Forward.

The relationship between the two is therefore one of governance and execution. A programme may determine that a specific capability area requires exercising, that particular sectors should participate, that defined metadata and artefacts must be generated, and that certain findings require re-testing. The lifecycle then provides the process through which these requirements are translated into exercise objectives, scope, scenario design, preparation, execution, evidence collection, evaluation, and follow-up. In this way, the lifecycle acquires institutional relevance through the programme, while the programme acquires repeatable operational form through the lifecycle.

The same relationship also establishes the foundation for Part III. The dissertation does not examine artificial intelligence as an isolated technological capability or as an unrestricted general-purpose assistant for exercise planning. Instead, AI is introduced into a governance environment in which required inputs, expected outputs, decision responsibilities, information-access conditions, artefact structures, and evidence expectations have already been defined. This is methodologically important because the evaluation of AI support depends on the existence of an external standard against which its outputs can be assessed. Without such a reference structure, claims concerning improved quality, completeness, consistency, or traceability would remain largely subjective.

The programme concept therefore provides the **reference frame for evaluating AI-supported outputs**. The Cyber Exercise Taxonomy and Minimum Dataset define the information that must be captured. The programme's artefact logic defines which products must exist and how they should relate to one another. The measurement framework defines how programme objectives can be translated into questions and indicators. The information-governance model defines what data may be processed, shared, or retained. The after-action and re-testing mechanisms define how exercise evidence must contribute to improvement. These requirements subsequently become design constraints for the AI-enhanced lifecycle and for the Auditable Agentic Workflow developed in Part III.

The programme concept also provides the accountability context for AI integration. The dissertation's later emphasis on human authority, provenance, bounded AI use, source grounding, phase-gated progression, and human approval does not emerge independently from the AI component of the research. These requirements follow logically from the governance principles established at national level. If the national programme requires traceable evidence, clearly assigned responsibility, controlled information access, and auditable improvement decisions, then AI-supported exercise processes must preserve those same properties. AI integration is therefore subordinate to the governance architecture rather than the reverse.

This relationship is particularly relevant in regulated, defence-oriented, and accountability-sensitive environments. In such contexts, a technologically capable system is not sufficient if its outputs cannot be traced to approved requirements, if the information it uses exceeds authorized access, or if responsibility for consequential decisions becomes unclear. The programme concept therefore provides the institutional boundary conditions within which AI augmentation can be considered acceptable. It allows the dissertation to examine not simply whether AI can generate useful exercise content, but whether AI-supported processes can operate within a governance system that requires evidence, accountability, confidentiality, and human decision authority.

The programme concept further strengthens the cumulative logic of the dissertation by creating a clear distinction between the three levels of contribution. The European Convergence Framework developed in Part I identifies **what is converging** and establishes the external baseline. The National Cyber Exercise Governance Model developed in the present section determines **how this baseline can be translated into a national capability system**. The ENISA-based AI-Enhanced Lifecycle subsequently defines **where AI support can be introduced within a recognized exercise process**, and the Auditable Agentic Workflow in Part III determines **how that support can be operationalized through bounded, traceable, and human-controlled processes**.

Accordingly, the national programme concept should be understood as the central connective mechanism of the dissertation rather than as an isolated practical recommendation. It links European convergence to national implementation, national strategy to exercise activity, exercise activity to evidence, evidence to corrective action, and governance requirements to controlled AI operationalization. Its scientific function is therefore to provide the institutional and methodological context within which the later lifecycle and workflow contributions can be evaluated coherently.

The programme concept also clarifies the boundaries of the dissertation's claims. It does not imply that a single programme architecture should be transferred unchanged to every country. National constitutional arrangements, cybersecurity authorities, sectoral responsibilities, security classifications, and exercise maturity differ considerably. The proposed model therefore specifies functional governance requirements rather than one mandatory institutional configuration. Similarly, the model does not claim that programme-level indicators can be reduced to a universal national resilience score. Its purpose is to create a more structured and auditable evidence base through which capability development can be observed, governed, and compared over time.

The cumulative architecture of the dissertation can therefore be expressed conceptually as a progression from **European convergence**, through **national governance and comparability**, to **lifecycle-level operationalization** and finally **auditable organizational AI support**. Within this

progression, the national programme is the mechanism that converts the analytical findings of Part I into a governable system and provides the control environment required for the contributions developed in Sections 2.5 and Part III. This positioning is essential to the overall research argument: responsible AI-supported cyber exercising is only meaningful when it operates within a system that already defines strategic objectives, governance responsibilities, information requirements, evidence expectations, and mechanisms for verified improvement.

2.5 ENISA-Based AI-Enhanced Cyber Exercise Lifecycle[104]

The National Cyber Exercise Governance Model developed in Section 2.4 establishes the institutional, programme-level, informational, and improvement-oriented conditions required for cyber exercises to function as a national resilience instrument. It defines how strategic priorities should be translated into exercise objectives, how exercise portfolios should be governed, how common artefacts and datasets can support comparability, and how findings should be connected to corrective action and re-testing. However, a governance model alone does not determine how an individual cyber exercise should progress from an initial requirement to a completed exercise, evaluated findings, corrective actions, and subsequent verification. A common lifecycle is therefore required to translate programme-level governance into repeatable exercise practice.

This dissertation addresses that requirement through an **ENISA-based AI-Enhanced Cyber Exercise Lifecycle**. The lifecycle adopts the ENISA Cybersecurity Exercise Methodology as its methodological backbone and extends it with a controlled artificial intelligence support layer. The purpose is not to replace ENISA with a new exercise methodology, nor to claim novelty by redefining lifecycle phases that are already well established in mature exercise practice. Instead, the contribution lies in specifying where AI can support recognized exercise activities, under which governance conditions such support should operate, which outputs should remain subject to human approval, and how evidence and provenance should be preserved across the lifecycle.

The lifecycle follows the six principal phases defined by ENISA: **Initiation, Design, Preparation, Execution, Evaluation, and Moving Forward**. These phases remain conceptually unchanged. The dissertation adds operational granularity by linking each phase to defined outputs, human decision points, evidence requirements, information-governance constraints, and potential AI augmentation. In this way, the lifecycle functions as a methodological interface between the National Cyber Exercise Governance Model and the organizational implementation developed later in Part III.

This positioning is central to the three-layer architecture of the dissertation. Part I establishes the international and European convergence baseline for credible and auditable cyber exercises. Part II translates that baseline into national governance, common programme structures, comparable data, measurement logic, and a structured lifecycle. Part III subsequently operationalizes the lifecycle-level AI augmentation through an Auditable Agentic Workflow. The present chapter therefore occupies an intermediate position. It defines **where** AI can support the exercise process and **under which controls**, while Part III defines **how** these augmentation opportunities are implemented technically and organizationally.

The lifecycle is governed by a simple but consequential principle: **AI may recommend, structure, analyze, and assist, but accountable humans retain decision authority**. Artificial intelligence may support requirement analysis, objective development, scenario engineering, inject preparation, evidence organization, consistency checking, evaluation, and reporting. It should not autonomously determine the strategic purpose of the exercise, approve scenarios, release high-impact injects, make authoritative legal interpretations, determine final evaluation findings, or

[104] The ENISA-based AI-Enhanced Cyber Exercise Lifecycle is not a standalone scientific contribution, but rather the methodological link between C2 and C3.

assign binding remediation priorities. Human responsibility therefore remains explicit across all phases.

The scientific contribution of this lifecycle is consequently not the introduction of AI into exercise planning as such. Generative AI can already be used informally to draft text, propose scenarios, or summarize information. The contribution lies in embedding AI support into a recognized exercise methodology in a manner that preserves governance, traceability, proportionality, information security, accountability, and lifecycle continuity. This distinction is particularly important in defence-oriented and regulated environments, where the quality of a generated artefact is not sufficient if its origin, assumptions, approval status, and downstream use cannot be reconstructed.

State of the Art of Exercise Lifecycles

The state-of-the-art analysis conducted for this dissertation demonstrates that mature exercise practice increasingly follows a lifecycle-oriented rather than event-oriented logic. An exercise is not treated simply as the period during which participants react to a simulated incident. Instead, it is embedded in a structured sequence connecting strategic intent, objectives, design, preparation, controlled execution, evaluation, corrective action, and subsequent improvement.

Several established approaches were examined in the dissertation, including the **Homeland Security Exercise and Evaluation Program (HSEEP), BSI- and PDCA-oriented approaches, the German LÜKEX exercise cycle, the process-based cyber exercise approach associated with Skopik and Leitner, and the ENISA Cybersecurity Exercise Methodology**. Although these approaches differ in terminology, institutional origin, scope, and intended users, they exhibit important methodological convergence.

The first common characteristic is **objective orientation**. Mature exercise methodologies begin by identifying why an exercise is needed and which capabilities, processes, decisions, or interfaces it is expected to validate. Objectives are therefore not an administrative addition to the exercise. They determine scope, scenario design, participant selection, evaluation criteria, and ultimately whether meaningful conclusions can be drawn from the exercise.

The second shared characteristic is **structured preparation**. Mature methodologies distinguish between the decision to conduct an exercise and the detailed preparation required to make that exercise useful. This includes role definition, scenario development, exercise control, logistics, participant preparation, communication planning, evaluation arrangements, and, where relevant, technical environment preparation.

The third characteristic is **controlled execution**. Exercise play must remain sufficiently realistic to generate meaningful participant behaviour while remaining bounded by predefined rules, safety considerations, exercise objectives, and facilitator control. This is particularly important in cyber exercises, where technical actions may interact with production-like systems, sensitive data, external organizations, or critical service processes.

A fourth common characteristic is **evaluation as an integral lifecycle function**. Evaluation is not treated as an activity that begins after the exercise has finished. Evidence requirements must already be considered during design and preparation. Evaluators need to know which behaviors, decisions, technical actions, timelines, and outcomes are relevant to the stated objectives. The exercise therefore becomes both a performance event and a structured data-collection opportunity.

The fifth shared characteristic is **continuous improvement**. Exercise value is realized only when observations are converted into findings, findings into corrective actions, and corrective actions into verified improvement. Mature models therefore close the lifecycle through after-action reporting, improvement planning, ownership assignment, follow-up, and re-testing.

Within these common principles, the individual methodologies contribute different strengths.

HSEEP provides a particularly strong programme-management and standardization perspective. It connects exercise programme management, design and development, conduct, evaluation, and improvement planning within a continuous cycle. Its emphasis on structured artefacts, After-Action Reports, Improvement Plans, and corrective-action tracking is particularly relevant for the dissertation's focus on repeatability and evidence-oriented exercise governance.

BSI- and PDCA-oriented approaches contribute a management-system perspective. Exercises are treated as recurring validation mechanisms within wider emergency-management, continuity, and resilience processes. The Plan–Do–Check–Act logic reinforces the idea that preparedness must be periodically tested, assessed, corrected, and tested again rather than treated as a static organizational state.

LÜKEX demonstrates how lifecycle thinking can be applied at national and inter-organizational level. It highlights extensive preparation, multi-stakeholder coordination, strategic crisis-management execution, structured evaluation, and cross-cycle learning. This is particularly relevant to national cyber resilience because weaknesses often occur at institutional interfaces rather than within one technical team.

The **process-based cyber exercise approach** contributes a more explicitly cyber-specific perspective. It highlights that scenario progression, inject sequencing, participant interaction, observable reactions, and evaluation requirements are interconnected. In this sense, the lifecycle is not merely chronological; it also represents dependencies between exercise artefacts.

Finally, the **ENISA Cybersecurity Exercise Methodology** provides the most directly relevant European cybersecurity-specific lifecycle examined in this dissertation. It structures exercise development into six phases - Initiation, Design, Preparation, Execution, Evaluation, and Moving Forward - and supports these phases through templates, checkpoints, and explicit progression logic.

The state-of-the-art analysis therefore produces two important findings. The state-of-the-art analysis therefore produces two important conclusions. First, there is no scientific need to create a completely new exercise lifecycle merely to demonstrate originality. The principal stages of mature exercise methodology are already well established and show considerable convergence. Second, an important gap remains between high-level lifecycle guidance and the requirements of scalable, comparable, evidence-oriented, and AI-supported national exercise programmes. Existing methodologies do not systematically define how AI support should be bounded, where human approval should be mandatory, what evidence of AI-supported decisions should be retained, or how provenance should be preserved across lifecycle artefacts.

The contribution developed in this chapter therefore adopts a recognised lifecycle rather than replacing it and concentrates scientific originality on **governance-compatible augmentation**.

Selection of ENISA as the Baseline

The ENISA Cybersecurity Exercise Methodology is selected as the primary methodological baseline for the AI-enhanced lifecycle. This choice follows from the comparative analysis rather than from institutional preference.

A first reason is its cybersecurity-specific scope. Other exercise methodologies provide highly valuable principles, but several originate from emergency management, civil protection, business continuity, or broader crisis-management environments. ENISA is explicitly focused on cybersecurity exercises and therefore addresses the central research object of this dissertation more directly.

A second reason is its complete end-to-end structure. The six phases provide a sufficiently comprehensive process from initial purpose to post-exercise improvement. This structure aligns well with the National Cyber Exercise Governance Model developed in Section 2.4, because

programme-level requirements can be translated into phase-specific exercise outputs and subsequently into evidence and improvement activities.

A third reason is its European institutional relevance. Since Part I of the dissertation identifies emerging European convergence around lifecycle orientation, governance, evidence, documentation, and improvement, the use of ENISA as the baseline strengthens conceptual continuity across the dissertation. The methodology provides a recognized European process architecture that can be embedded within national governance without requiring the dissertation to introduce an entirely separate lifecycle.

A fourth reason is its suitability for augmentation. ENISA's methodology is sufficiently structured to support explicit phase transitions and sufficiently flexible to accommodate different exercise types, sectors, scales, and maturity levels. This makes it suitable for the addition of AI support without constraining the methodology to one technological environment.

At the same time, the selection of ENISA does not imply that the methodology is complete for the purposes of this research. Its intentional generality means that it does not prescribe in detail how AI augmentation should be governed, how provenance should be retained across AI-supported artefacts, which human approval points should apply, or how lifecycle evidence should connect to the national programme's taxonomy, minimum dataset, and measurement framework. The scientific extension proposed in this dissertation therefore concentrates precisely on these areas. The relationship between the three levels can be stated clearly. **ENISA defines the exercise process. The AI-enhanced lifecycle defines the permitted augmentation points and governance requirements. The Auditable Agentic Workflow developed in Part III subsequently defines how those augmentation points can be operationalized.**

AI Augmentation Principles

The AI-enhanced lifecycle is governed by a set of cross-cutting principles that apply across all six phases. These principles are necessary because the value of AI support depends not only on technical capability but also on whether its use remains controllable, traceable, proportionate, legally compatible, and appropriate to the sensitivity of the exercise environment. The increasing use of generative AI in cybersecurity creates opportunities for automation, analysis, content generation, and decision support, while simultaneously introducing new risks related to reliability, misuse, data handling, and human dependence (Bányász et al., 2024).[105] The AI-enhanced lifecycle developed in this dissertation therefore treats AI augmentation as a governed capability rather than as unrestricted automation.

The first principle is **human authority**. Human decision authority must remain intact throughout the lifecycle. AI can assist in extracting requirements, identifying inconsistencies, proposing objectives, developing scenario variants, structuring injects, summarizing evidence, clustering observations, and drafting reports. However, these outputs remain recommendations or analytical support until they have been reviewed and accepted by an accountable human role. This is particularly important in defence-oriented and regulated environments, where technically plausible outputs may nevertheless be strategically, legally, politically, or operationally inappropriate.

The second principle is **traceability and provenance**. AI-supported artefacts must remain traceable to both their source material and their transformation history. Provenance should allow later reconstruction of which material informed an output, which component produced it, what assumptions were introduced, which version was generated, what changes were made, who reviewed the output, and whether it was accepted, revised, or rejected. Provenance is therefore treated as a lifecycle property rather than as a reporting feature added retrospectively.

[105] Bányász, P., Szádeczky, T., & Váczi, K. B. (2024). “The Relationship Between Generative Artificial Intelligence and Cybersecurity.” In Proceedings of the Central and Eastern European eDem and eGov Days 2024 (pp. 209–215). ACM.

The third principle is **bounded AI use**. AI should operate within predefined purposes, permitted inputs, expected outputs, and prohibited actions. Its access should be limited by data classification, exercise objectives, sectoral constraints, approved threat information, legal restrictions, output format, and permitted automation level. This distinguishes controlled augmentation from unrestricted use of a general-purpose AI assistant. Bounded use also reinforces proportionality, because a small organizational tabletop exercise and a national multi-sector exercise should not automatically employ the same level of AI support.

The fourth principle is **phase-gated progression**. The lifecycle should prevent AI-supported artefacts from automatically propagating from one phase to the next simply because they have been generated. Critical outputs should be reviewed before they become authoritative inputs to subsequent activities. Requirements identified during Initiation should be validated before they shape objectives; objectives should be approved before they drive scenario design; scenarios and injects should be reviewed before execution; AI-assisted observations should be validated before they become findings; and findings should be approved before they are converted into corrective actions. This creates a controlled sequence of production, review, validation, approval, and progression.

These four primary principles are complemented by **data minimization, security, proportionality, and accountability**. Cyber exercises may involve sensitive information concerning vulnerabilities, critical infrastructure, internal crisis procedures, threat intelligence, staff responsibilities, or national-security-relevant dependencies. The availability of such information does not justify unrestricted processing. AI-supported tasks should therefore use only the data necessary for the relevant purpose, under approved access controls and environments. Accountability must remain attributable to a defined human role or governance body for every significant lifecycle artefact. AI-generated content should therefore remain a draft, proposal, or analytical aid until formal human acceptance.
Together, these principles establish the governance boundary within which AI can support the ENISA lifecycle.

AI Support Across the Six Lifecycle Phases

The AI-enhanced lifecycle introduces differentiated support across each ENISA phase. AI is not applied uniformly. The type and degree of augmentation depend on the purpose of the phase, the sensitivity of its outputs, and the level of judgement involved.

A useful analytical structure for every phase is:

ENISA purpose → potential AI enhancement → required human control → principal output.

2.5.1.1 *Phase 1 – Initiation*

The Initiation phase determines why the exercise should take place and establishes its basic feasibility, strategic relevance, and governance context.

At this stage, AI can support the analysis of large volumes of input material. Relevant sources may include national strategies, organizational policies, regulatory requirements, risk assessments, previous After-Action Reports, threat assessments, continuity plans, and programme-level priorities.

Potential AI-supported activities include:

- extraction of relevant requirements;
- classification of identified requirements;
- identification of recurring weaknesses from previous exercises;
- stakeholder clustering;

- initial mapping of responsibilities;
- identification of relevant sectors and dependencies;
- preliminary scope suggestions;
- identification of constraints;
- comparison with previous exercise objectives;
- and identification of potential gaps in proposed exercise coverage.

AI can therefore reduce the manual effort required to transform heterogeneous input material into a structured initiation package.

However, several decisions remain explicitly human.

Human authorities determine:

- whether an exercise is required;
- its strategic purpose;
- acceptable scope;
- participating organizations;
- confidentiality level;
- major exclusions;
- resource allocation;
- and whether the exercise proceeds to Design.

The main AI-supported output of Initiation is a **structured requirements and scope package** rather than an automatically approved exercise mandate.

The governance gate asks:

Is the proposed purpose, scope, stakeholder structure, and requirement basis sufficiently justified and approved to proceed to Design?

2.5.1.2 *Phase 2 – Design*

The Design phase translates strategic purpose into measurable objectives and a coherent exercise concept.

AI can support:

- refinement of exercise objectives;
- conversion of broad aims into SMART objectives;
- identification of relevant capability areas;
- proposal of KPIs or observable criteria;
- mapping between regulatory requirements and exercise objectives;
- mapping between objectives and stakeholder roles;
- scenario-theme generation;
- alternative scenario branches;
- initial dependency mapping;
- identification of potential evidence sources;
- and consistency checks between scope, objectives, actors, and expected outputs.

The main value of AI in Design is not unrestricted creativity but **structured consistency**.

For example, an AI-supported system can identify that an objective relating to third-party incident coordination has no supplier or provider represented in the proposed stakeholder structure. It can also identify that a scenario theme does not generate any observable event related to a stated continuity objective.

Human planners remain responsible for approving:

- objectives;
- KPIs;
- capability targets;
- scenario direction;
- sensitive threat assumptions;
- participant roles;
- exercise realism;
- and legal or policy interpretation.

The principal output is an **approved exercise design package** containing purpose, scope, objectives, evaluation logic, stakeholder mapping, and scenario concept.

The governance gate asks:

Are the objectives measurable, the scope appropriate, the scenario concept credible, and the evidence requirements sufficiently defined to proceed to Preparation?

2.5.1.3 *Phase 3 – Preparation*

Preparation converts the approved design into executable exercise artefacts.

This phase represents one of the areas with the greatest potential for AI augmentation because cyber exercise preparation is typically labour intensive and requires consistency across many interdependent artefacts.

Potential AI support includes:

- development of the scenario backbone;
- generation of scenario branches;
- construction of MSEL structures;
- drafting of injects;
- creation of alternative inject variants;
- consistency checking across injects;
- role-specific briefing material;
- participant documentation;
- facilitator notes;
- evaluator guidance;
- communication templates;
- simulated media content;
- preparation of evidence-capture structures;
- and mapping of injects to objectives and evaluation criteria.

A key requirement is that AI-generated content remains connected to the approved design.

Each significant inject should therefore be traceable to:

objective → scenario event → expected participant action → observable evidence → evaluation criterion.

This relationship is central to the dissertation's evidence-oriented approach.

Human exercise planners and facilitators remain responsible for:

- scenario plausibility;

- escalation intensity;
- legal and ethical acceptability;
- technical safety;
- sensitive information handling;
- timing;
- participant realism;
- and final approval of the MSEL and exercise package.

The main output is an **approved execution package**, including scenario artefacts, MSEL/inject structure, participant materials, control documentation, and evaluation arrangements.

The governance gate asks:

Is the exercise operationally, technically, legally, and organizationally ready to run?

2.5.1.4 *Phase 4 – Execution*

Execution is the most sensitive phase for AI augmentation because the exercise is actively influencing participant behaviour.

AI support should therefore be more constrained than during planning.

Potential functions include:

- monitoring exercise progression;
- supporting facilitator situational awareness;
- comparing actual progress with the planned MSEL;
- identifying missed triggers;
- identifying timing anomalies;
- suggesting bounded inject adaptation;
- structuring incoming observations;
- categorising exercise events;
- assisting with evidence tagging;
- and supporting exercise-control information management.

AI may be particularly useful in large-scale or multi-organizational exercises where control teams receive substantial amounts of information simultaneously.

However, autonomous scenario modification should generally be avoided.

The human exercise director or controller should retain authority over:

- release of injects;
- changes in scenario direction;
- escalation;
- suspension;
- safety interventions;
- acceptance of deviations;
- and any action affecting participants.

AI may suggest that an inject should be delayed or adapted, but the controller decides whether this occurs.

The main output of Execution is therefore not simply completed scenario play, but a **structured exercise record** containing decisions, events, participant responses, observations, timestamps, and supporting evidence.

The execution governance principle is:

2.5.1.5 *AI may support awareness and orchestration; humans retain control of exercise play.*

2.5.1.6 *Phase 5 – Evaluation*

The Evaluation phase converts observations and evidence into findings.

Potential AI support includes:

- consolidation of evaluator notes;
- clustering of observations;
- de-duplication of similar findings;
- timeline reconstruction;
- mapping evidence to objectives;
- comparison against predefined criteria;
- identification of inconsistent evidence;
- generation of draft finding statements;
- summarisation of participant feedback;
- comparison with previous exercise cycles;
- and drafting sections of the After-Action Report.

AI can significantly reduce administrative effort in this phase, particularly when large exercises generate substantial quantities of qualitative material. However, evaluation contains judgement-intensive decisions.

AI-generated findings should never become authoritative solely because they are statistically frequent or linguistically convincing.

Human evaluators determine:

- whether evidence is sufficient;
- whether an observation represents a genuine capability gap;
- contextual significance;
- severity;
- causality;
- attribution;
- and final evaluative conclusions.

The required traceability relationship is:

objective → evidence → observation → finding → conclusion.

The main output is an **approved evaluation package and After-Action Report**.

The governance gate asks:

Are the findings supported by sufficient evidence, appropriately contextualised, and approved by responsible evaluators?

2.5.1.7 *Phase 6 – Moving Forward*

The Moving Forward phase transforms evaluation into capability improvement.

ENISA emphasises that the value of an exercise depends on whether identified weaknesses lead to action and whether those actions are subsequently verified.

AI can support:

- prioritisation proposals;
- grouping of corrective actions;
- mapping findings to responsible capability areas;
- identification of recurring weaknesses;
- preparation of remediation drafts;
- comparison with previous improvement plans;
- proposal of follow-up indicators;
- identification of suitable re-test scenarios;
- generation of management summaries;
- and preparation of differentiated dissemination products.

AI may also support longitudinal programme-level analysis by identifying patterns across multiple exercises.

Human authorities remain responsible for:

- accepting findings;
- defining remediation priorities;
- assigning ownership;
- allocating resources;
- accepting residual risk;
- deciding which issues require escalation;
- and determining when and how re-testing should occur.

The main output is a **validated improvement and re-testing plan**.

The lifecycle then returns to a new Initiation phase, creating the continuous capability-development cycle:

exercise → evidence → evaluation → improvement → verification → new exercise.

ENISA phase	Primary purpose	Potential AI support	Required human control	Main output
Initiation	Establish need, purpose, feasibility and scope	Requirement extraction, stakeholder mapping, prior-finding analysis, scope suggestions	Strategic purpose, scope, participation, confidentiality and authorisation	Approved requirements and scope package
Design	Define objectives, capabilities and exercise concept	Objective refinement, KPI suggestions, scenario options, dependency and evidence mapping	Approval of objectives, KPIs, scenario direction and assumptions	Approved exercise design
Preparation	Produce executable exercise artefacts	Scenario backbone, MSEL, injects, briefing material, consistency checks	Scenario realism, safety, inject approval and readiness decision	Approved execution package
Execution	Conduct controlled exercise play	Monitoring, evidence tagging, pacing analysis, bounded adaptation suggestions	Inject release, scenario changes, escalation, safety and control	Structured exercise record
Evaluation	Analyse evidence and derive findings	Evidence organisation, clustering, objective mapping, draft AAR	Validation of evidence, findings,	Approved AAR and evaluation package

			severity and conclusions	
Moving Forward	Convert findings into improvement	Corrective-action structuring, trend analysis, re-test suggestions, dissemination drafts	Priority, ownership, resource decisions and acceptance	Improvement and re-testing plan

TABLE 13 - AI AUGMENTATION ACROSS THE ENISA LIFECYCLE

The table illustrates an important property of the model: **AI authority does not increase uniformly as the lifecycle progresses**. The most suitable AI activities are generally those involving structured analysis, consistency checking, repetitive artefact production, and information organisation. Activities requiring strategic judgement, legal responsibility, exercise-control authority, or formal evaluation remain human-led.

Human Approval and Governance Gates

Human approval gates are a central governance mechanism of the proposed lifecycle. This governance logic is consistent with the principle that AI trustworthiness depends not solely on model performance, but also on the ability to constrain, supervise and intervene in AI-supported processes[106].

The concept builds on the progression logic already present in the ENISA methodology but extends it explicitly to AI-supported artefacts. The purpose is to prevent unreviewed AI output from becoming an authoritative input for subsequent lifecycle phases.

Without such gates, errors can propagate.

For example, an incorrectly extracted regulatory requirement in Initiation may produce an inappropriate objective in Design. That objective may then generate irrelevant injects in Preparation, which may create misleading observations during Execution and ultimately produce invalid evaluation findings.

AI-supported lifecycle governance must therefore prevent **error propagation across artefact dependencies**.

The dissertation introduces five principal transition gates.

2.5.1.8 *Governance Gates*

The proposed AI-enhanced cyber exercise lifecycle incorporates five mandatory transition gates to ensure that each phase is completed, reviewed, and formally approved before the exercise progresses. These gates establish a structured governance mechanism through which exercise quality, methodological consistency, operational readiness, and accountability are maintained. Although AI-supported components may assist with the preparation and verification of exercise artefacts, final decisions remain the responsibility of authorised human stakeholders.

Gate 1 – Initiation to Design

The first transition gate establishes whether the exercise has a sufficiently defined and justified strategic foundation. Reviewers assess the exercise purpose, underlying requirements, scope, stakeholders, operational constraints, assumptions, information sensitivity, and overall feasibility. Particular attention is given to ensuring that the proposed exercise addresses relevant organisational or national preparedness objectives. AI-generated requirements analyses and

[106] Turaga, V. S. P., Páhi, T., Tjoa, S., Siami-Namini, S., & Siami Namin, A. (2025). CO2 (Co-Compliance Officer): An LLM-based ontology-driven methodology for generating knowledge graphs and AI compliance checking. IEEE Access, 13, 210576–210606.

scope recommendations remain provisional until reviewed and approved. Based on this assessment, the responsible authority may approve progression to the Design phase, request revisions, or terminate the proposed exercise.

Gate 2 – Design to Preparation

The second gate evaluates the coherence and completeness of the exercise design before detailed preparation begins. The review focuses on the alignment between exercise objectives, targeted capabilities, performance indicators, scenario assumptions, stakeholder responsibilities, organisational dependencies, and planned evaluation evidence. Its primary purpose is to prevent incomplete or contradictory design elements from influencing subsequent exercise development. Progression is authorised only when the proposed design provides a consistent foundation for scenario elaboration, inject development, and evaluation planning. The decision is either to approve the design or require revisions.

Gate 3 – Preparation to Execution

The third gate constitutes the formal exercise-readiness assessment and represents a particularly important governance checkpoint. Reviewers determine whether the scenario and Master Scenario Events List (MSEL) are sufficiently complete, whether injects are coherent and relevant, and whether the technical environment, evaluation arrangements, participants, facilitators, and exercise-control structures are prepared. Safety requirements, confidentiality measures, and contingency procedures must also be verified. This assessment is especially important where AI-supported components have contributed substantially to scenario development or inject generation, as their outputs require validation for operational relevance, accuracy, and consistency. The responsible authority may approve execution, postpone the exercise, request corrective action, or cancel it if essential readiness conditions cannot be satisfied.

Gate 4 – Execution to Evaluation

The fourth transition gate determines whether the evidence collected during exercise execution is sufficiently complete and reliable to support a defensible evaluation. This assessment considers observation records, event logs, inject delivery histories, evaluator notes, participant feedback, relevant technical evidence, and documented deviations from the planned scenario. The objective is not merely to confirm that documentation exists, but to determine whether the available evidence adequately supports subsequent analysis of participant performance and exercise objectives. AI-assisted tools may contribute to evidence consolidation, structuring, and completeness checks; however, human evaluators retain responsibility for determining evidentiary adequacy. Where significant gaps are identified, additional clarification or evidence reconstruction may be required before formal evaluation proceeds.

Gate 5 – Evaluation to Moving Forward

The final transition gate validates the evaluation findings before they are translated into formal improvement measures. Reviewers examine whether the identified weaknesses and strengths are adequately supported by evidence, whether their severity and operational significance have been interpreted appropriately, and whether the proposed corrective actions are relevant and assignable to responsible stakeholders. Findings may be approved, revised, consolidated, or rejected based on this assessment. Only validated findings are transferred into the improvement process, where they can be associated with responsible owners, implementation deadlines, verification criteria, and subsequent re-testing. This gate ensures that the exercise contributes to measurable capability development rather than concluding with an unverified list of lessons identified.

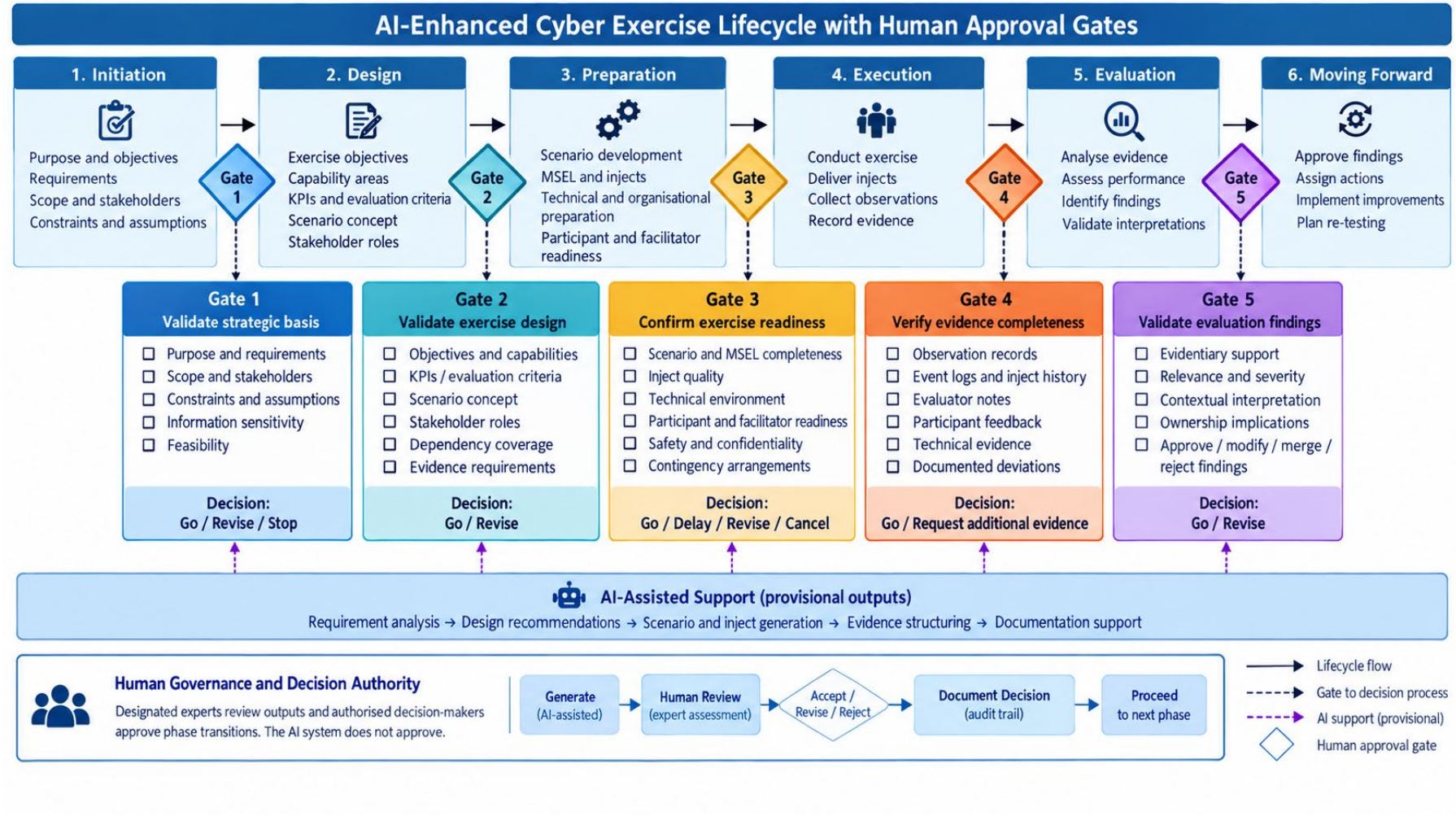


FIGURE 22 - HUMAN APPROVAL GATES

The figure illustrates how five mandatory human approval gates control progression across the six phases of the ENISA-based AI-Enhanced Cyber Exercise Lifecycle. Each gate verifies the completeness, quality, and readiness of exercise artefacts before authorised stakeholders approve progression, request revisions, or require additional evidence. AI-supported components assist with analysis, planning, and documentation, while human decision-makers retain approval authority and ensure traceability and accountability.

2.5.1.9 *Governance Roles, Decision Authority and Human Revision*

The institutional arrangements supporting these transition gates depend on the scale, purpose, and complexity of the exercise. At the organisational level, decisions may be taken by an exercise planning committee or designated exercise authority. National or defence-oriented exercises may require coordination among governmental authorities, military and civilian stakeholders, sector representatives, legal and compliance functions, technical experts, and exercise-control personnel. Consequently, the proposed model does not prescribe a universal institutional structure but requires clearly assigned responsibilities and documented decision authority.

For each transition gate, three governance responsibilities must be explicitly defined: the stakeholder responsible for proposing the artefact or phase transition, the stakeholder responsible for reviewing its adequacy, and the authority empowered to approve progression. These functions may be distributed across different individuals or organisational bodies depending on the exercise context. Regardless of the arrangement, AI-supported components may generate recommendations and supporting evidence but cannot independently approve artefacts, authorise phase transitions, or assume accountability for consequential decisions.

The rejection of an exercise artefact does not necessarily require termination of the workflow. Instead, the model distinguishes between direct human revision and bounded AI-supported regeneration. In the first case, the responsible expert corrects or supplements the artefact. In the second, an AI component revises its output based on explicit reviewer feedback, while maintaining the original task boundaries, approved information sources, and applicable governance constraints.

This approach enables iterative improvement without permitting uncontrolled AI generation or bypassing mandatory approval requirements. Each revision must remain traceable through documented reviewer decisions, artefact versions, and, where applicable, updated source references. An artefact can advance to the next lifecycle phase only after receiving the required human approval.

The resulting governance sequence is expressed as:

Generate → Human Review → Accept / Revise / Reject → Document Decision → Proceed

Together, the five transition gates, defined decision authorities, and controlled revision mechanisms establish a human-supervised governance structure for the AI-enhanced cyber exercise lifecycle. The model supports methodological consistency, operational accountability, and auditable progression while preserving the central principle that AI assists the process but does not replace professional judgement or human decision-making authority.

Evidence and Provenance Flows

Evidence and provenance are treated as central properties of the AI-enhanced lifecycle because repeatability and auditability require more than retaining final documents. It must be possible to understand how significant outputs were created and how they influenced subsequent decisions.
The lifecycle therefore establishes two related but distinct forms of traceability. The first is **exercise evidence traceability**. This connects programme requirements and exercise objectives to scenario elements, injects, participant behavior, observations, findings, corrective actions, and re-testing. The second is **AI provenance traceability**. This connects source material, AI-supported transformations, generated drafts, human review, approved artefacts, and downstream use.

The two flows intersect throughout the lifecycle. A regulatory requirement identified during Initiation may be transformed into an approved objective during Design. That objective may influence a scenario and specific injects during Preparation. Participant responses during Execution generate evidence. Evaluation maps that evidence back to the objective and produces a finding. Moving Forward translates the finding into corrective action and later re-testing. If AI has supported any of these transitions, its role should remain reconstructable.

This requirement is particularly important because an apparently correct final document may conceal unsupported intermediate assumptions. Provenance therefore allows reviewers to distinguish between information derived from authoritative sources, content generated through AI transformation, human modifications, and final approved decisions.

A mature provenance record should retain the relevant source identifier, output version, timestamp, producing component, significant assumptions, review status, reviewer identity or responsible role, and the relationship to the final artefact. The level of detail should remain proportionate to the sensitivity and importance of the task, but higher-risk or legally relevant outputs require stronger traceability.
Provenance also interacts directly with information governance. Not all source materials or evidence should be centralized or made available to every AI-supported component. Restricted operational evidence may remain under local control while only approved metadata or summaries are used at programme level. Provenance therefore does not imply unrestricted data aggregation. It means that within authorized boundaries, the origin and transformation history of relevant information remain auditable.

This logic connects the lifecycle directly to Section 2.3, where the Taxonomy and Minimum Dataset define the programme-level information structure, and to Section 2.4, where differentiated information-access levels and evidence governance are integrated into the National Cyber Exercise Governance Model.

Methodological Role of the ENISA-Based AI-Enhanced Lifecycle

The ENISA-based AI-Enhanced Cyber Exercise Lifecycle serves as a methodological bridge between the National Cyber Exercise Governance Model (C2) and the Auditable Agentic AI Workflow (C3). It is not classified as a separate scientific contribution.

Its novelty does not lie in the six lifecycle phases themselves. These are adopted from ENISA and remain methodologically attributable to the ENISA Cybersecurity Exercise Methodology. The contribution lies in the systematic definition of controlled AI augmentation across those phases, combined with explicit human authority, provenance, bounded use, information-governance requirements, and phase-gated progression.

This distinction is scientifically important because it prevents the dissertation from claiming originality where established methodology already exists. Instead, the research extends the recognized process architecture in areas that become increasingly important when AI is introduced into exercise planning and evaluation.

The lifecycle also provides a direct response to the broader research problem. The comparability gap identified in the dissertation requires structured and repeatable exercise processes. The improvement gap requires evidence to flow from objectives through evaluation into corrective action and re-testing. The responsible-scaling gap requires AI support to reduce manual effort without weakening accountability or traceability. The AI-enhanced lifecycle brings these concerns together at the methodological level.

Within Part II, the lifecycle serves as the operational bridge between the National Cyber Exercise Governance Model and the organizational AI implementation developed in Part III. The governance model defines the institutional environment, programme requirements, information structures, and improvement logic. The lifecycle defines the repeatable process through which these requirements are applied to individual exercises. The Agentic Workflow subsequently implements selected lifecycle activities through bounded AI-supported components.

The resulting relationship is therefore cumulative rather than competitive. ENISA provides the recognized exercise methodology; the dissertation adds governance-controlled AI augmentation; and Part III operationalizes that augmentation through an auditable workflow.

The scientific claim remains deliberately bounded. The lifecycle does not demonstrate that AI automatically improves cyber resilience, nor does it imply that every lifecycle activity should be automated. Its contribution is to provide a structured, governance-compatible model through which AI support can be introduced in a way that remains reviewable, attributable, proportionate, and compatible with human decision authority.

In this way, Section 2.5 completes the methodological component of the national research layer. It translates the programme architecture of Section 2.4 into a repeatable exercise process and establishes the conditions under which AI can support that process without undermining accountability (see the Overview in Figure 22).The next step, developed in Part III, is to determine how these lifecycle-level augmentation points can be implemented as a concrete Auditable Agentic Workflow.

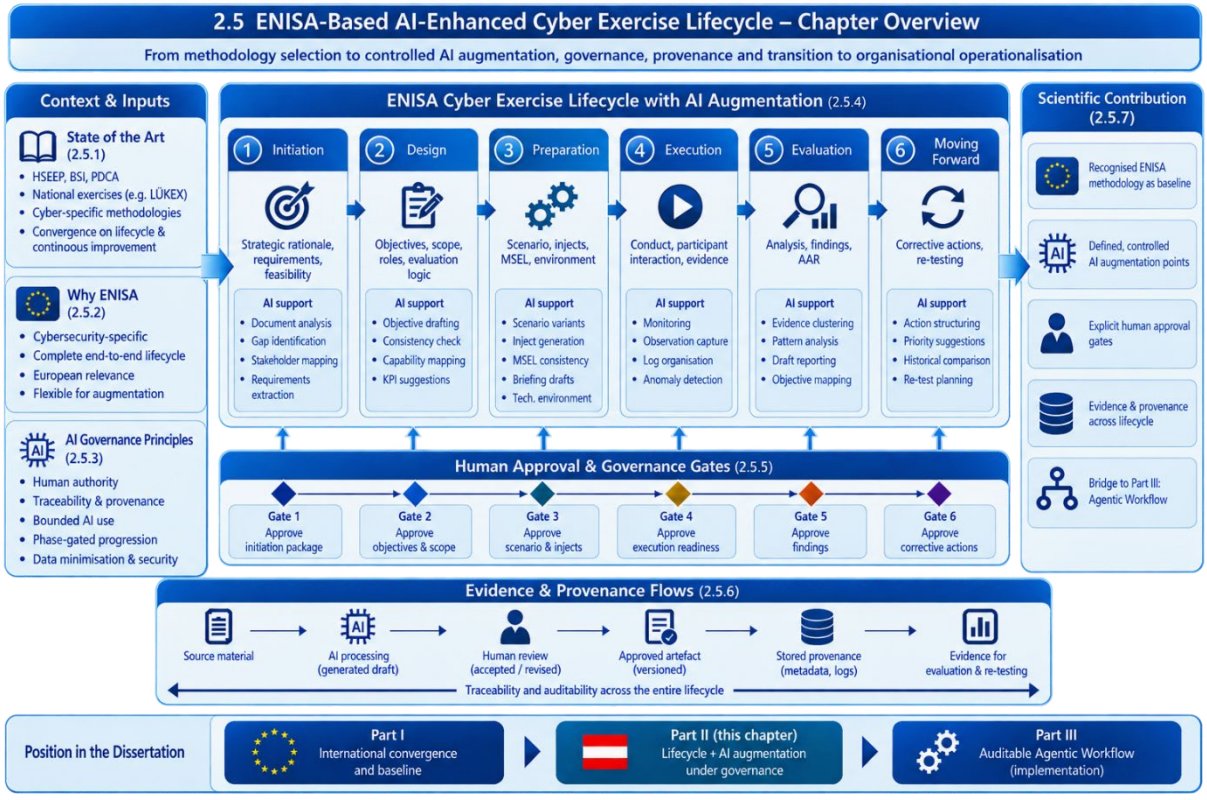


FIGURE 23 - METHODOLOGICAL OVERVIEW: ENISA-BASED AI-ENHANCED LIFECYCLE

2.6 Regulatory Pathway II – National Governance for Cyber Resilience

The national regulatory pathway explains how European cybersecurity and resilience requirements become national governance requirements and, subsequently, cyber exercise objectives. It is not a separate legal-compliance chapter; it is the translation mechanism connecting H1 with H2.
The pathway follows four stages: European requirement → Austrian legal and institutional implementation → national capability expectation → exercise-based validation. Different legal instruments create different forms of obligation, but many converge on operational capabilities such as risk management, incident response, continuity, recovery, dependency management, reporting, evidence, and management accountability.

NIS2 / NISG 2026

NIS2 is translated into Austria through NISG 2026 and the new national cybersecurity authority. For national exercise governance, the relevant capabilities include incident handling, escalation, reporting, continuity, crisis management, supply-chain security, effectiveness assessment, and coordination with national and European structures. Exercise scenarios can validate whether reporting information is available in time, whether crisis escalation works, whether supplier dependencies are managed, and whether multiple entities can contribute to a coherent national situation picture.

CER / RKEG and RKEV

The CER framework is implemented through the RKEG and RKEV. Its exercise relevance lies in the continuity of essential services, all-hazards resilience, dependency management, crisis coordination, and recovery. Cyber scenarios affecting critical entities should therefore test not only technical response but also whether essential functions can be preserved or restored and whether cross-sector consequences are managed.

DORA / DORA-VG

DORA creates explicit digital operational resilience testing requirements for the financial sector, implemented nationally through DORA-VG and supervisory arrangements. The national programme should recognise regulated testing as part of the wider resilience portfolio while avoiding duplication. Financial-sector exercises can also test interfaces with national cyber crisis management, external ICT providers, and cross-sector dependencies.

GDPR / Austrian Data Protection Framework

GDPR and Austrian data-protection governance influence how exercise participant data, technical evidence, communications, logs, and AI-supported artefacts are processed. Data minimization, purpose limitation, access control, retention, and lawful processing should therefore be built into exercise design and evidence management.

EU AI Act / Austrian AI Governance

The EU AI Act applies directly at European level, while Austrian structures support implementation, guidance, supervision, and coordination. For the national AI-enhanced lifecycle, relevant controls include human oversight, risk-based use, transparency, source grounding, data governance, security, and traceability. These requirements become concrete through human approval gates and provenance-aware artefact handling.

Implication for National Cyber Exercise Governance

National exercise governance should not create a separate exercise methodology for every legal regime. Instead, requirements should be mapped to a common capability and evidence structure. The source of an obligation may differ, while the exercise logic remains comparable: requirement → capability → objective → scenario → evidence → finding → corrective action → re-test.
This pathway allows regulation to shape exercise priorities without reducing cyber exercises to compliance events. The central purpose remains resilience improvement, while regulatory alignment provides an additional rationale for structured documentation, evidence, accountability, and follow-up.

2.7 Assessment of H2 and Part II Summary

Part II examined how the international and European baseline established in Part I can be translated into a coherent national cyber exercise governance system. The purpose of this research layer was not to determine whether national cyber exercises should exist, as such exercises are

already widely conducted, but to assess whether a structured national governance architecture can provide a more suitable basis for comparability, auditability, and resilience-oriented improvement than fragmented or ad hoc exercise activity. This question directly addresses H2 and MRQ2.
H2 proposed that a national defence-oriented cyber exercise ecosystem based on defined governance responsibilities, common quality criteria, comparable artefacts and datasets, coordinated exercise programmes, a structured lifecycle, and mandatory improvement loops provides a stronger framework for resilience-oriented capability development than isolated exercise activity. The findings of Part II support this proposition as a **constructive governance hypothesis**, although the dissertation deliberately avoids interpreting this support as universal empirical proof of increased national cyber resilience.

The first line of evidence derives from the **Austrian reference case** analyzed in Section 2.2. The case demonstrates that national cybersecurity and resilience are inherently distributed across governmental authorities, CSIRTs, regulators, critical entities, defence and security institutions, public administration, research organizations, private operators, and European cooperation structures. This institutional diversity confirms that national cyber resilience cannot be assessed adequately through isolated organizational exercises alone. Important capabilities exist at the interfaces between institutions, including escalation, information exchange, crisis coordination, public-private cooperation, cross-sector dependency management, and the translation of technical situational awareness into operational and strategic decision-making. The Austrian analysis therefore provides empirical support for the governance premise underlying H2: national exercise capability requires coordination across a distributed ecosystem rather than centralized ownership of all exercise activity.
At the same time, the Austrian reference case also shows that the existence of competent authorities, coordination bodies, CSIRTs, regulatory structures, and recurring exercise activities does not automatically create a comparable or continuously improving national exercise system. Institutional structures may support coordination during incidents, yet this does not in itself ensure that exercises use compatible information structures, generate comparable evidence, track recurring weaknesses, or verify whether corrective actions have been effective across exercise cycles. This distinction is important because H2 concerns not simply the existence of governance, but the quality and continuity of the governance mechanisms surrounding exercise activity.

Section 2.3 addressed this comparability problem through the **Cyber Exercise Taxonomy and Minimum Dataset.** The taxonomy provides a common classification structure for heterogeneous exercise activities, while the Minimum Dataset specifies the core information required for national portfolio visibility and longitudinal analysis. Their contribution lies in standardizing the description of exercises rather than standardizing the exercises themselves. This preserves legitimate diversity in exercise format while creating a common information layer through which national stakeholders can understand what was tested, by whom, at which capability level, under which conditions, and with which results.
The relevance of this information layer to H2 is substantial. Without common descriptors and minimum information requirements, exercise portfolios cannot be analyzed systematically and longitudinal comparison remains weak. The taxonomy and dataset therefore provide the informational precondition for national programme governance. They make it possible to identify capability coverage, under-exercised sectors, recurring findings, maturity patterns, and re-testing requirements without forcing all participants into one operational exercise model. The associated information-governance approach further demonstrates that national comparability does not require unrestricted centralization of sensitive operational evidence. Programme metadata, controlled artefacts, and restricted evidence can remain subject to differentiated access, classification, data minimization, and need-to-know principles.

Section 2.4 subsequently translated these comparability mechanisms into the **National Cyber Exercise Governance Model and Programme Blueprin**t, which constitutes the principal constructive contribution associated with H2. The model adds the programme layer required to connect strategic priorities with individual exercise activity. It introduces defined governance roles, portfolio management, standardized artefact logic, structured information requirements, KPI-oriented measurement, corrective-action ownership, and re-testing. In doing so, it shifts the analytical focus from the completion of exercises to the management of capability development across repeated exercise cycles.
The significance of this shift is central to the assessment of H2. An isolated exercise may identify weaknesses, but it cannot by itself demonstrate whether those weaknesses have subsequently been reduced. The programme model therefore treats the improvement loop as a constitutive element of national exercise governance. Exercise evidence is translated into findings; findings are assigned to responsible actors; corrective actions are implemented and monitored; and selected improvements are subsequently re-tested. The resilience value of the programme therefore lies not in exercise frequency alone, but in the continuity between evidence, remediation, and verification.

The **measurement framework** developed in Section 2.4.7 further supports this logic. The dissertation distinguishes between simple programme-activity metrics and evidence that is more directly related to capability development. The number of exercises, participants, or participating organizations may describe programme scale, but these indicators do not demonstrate resilience improvement. Stronger evidence is generated where objectives are connected to evaluation questions, findings are linked to corrective actions, remediation status is monitored, recurring weaknesses are identified, and improvement is verified through subsequent exercises or targeted re-testing. This distinction prevents the national governance model from reducing resilience to a single composite score and provides a more defensible basis for longitudinal assessment.

Section 2.5 adds a further methodological component by integrating the ENISA Cybersecurity Exercise Methodology into the national governance architecture and defining how AI support can be introduced under controlled conditions. ENISA provides the recognized six-phase lifecycle of Initiation, Design, Preparation, Execution, Evaluation, and Moving Forward. The dissertation does not claim these phases as original. Its contribution lies in linking them to the national programme architecture and specifying where AI augmentation may support lifecycle activities while preserving human authority, provenance, bounded use, data minimization, and phase-gated progression.
This lifecycle strengthens the assessment of H2 because a national governance model requires more than institutional roles and common datasets; it also requires a repeatable methodological process through which individual exercises can be planned, conducted, evaluated, and improved. The ENISA-based lifecycle provides this common process structure while remaining sufficiently flexible to support different exercise formats and maturity levels. It therefore contributes to consistency without imposing unnecessary methodological uniformity.

Section 2.6 complements the governance and lifecycle components through the national regulatory pathway. The analysis demonstrates that European and national requirements concerning cybersecurity governance, critical-entity resilience, operational resilience, data protection, and AI governance can be translated into exercise-relevant capabilities without reducing exercises to compliance events. The source of an obligation may vary, but the exercise logic remains structurally comparable: a requirement can be interpreted as a capability expectation, translated into an exercise objective, tested through an appropriate scenario, observed through evidence, and connected to corrective action and re-testing. This makes regulation relevant to exercise governance while preserving the distinction between formal legal compliance and operational validation.

Taken together, these findings support H2 because the proposed national governance architecture addresses the two principal weaknesses of fragmented exercise practice: lack of comparability and lack of verified improvement. The Cyber Exercise Taxonomy and Minimum Dataset create a common descriptive and information layer; the National Cyber Exercise Governance Model establishes institutional and programme-level coordination; the measurement framework connects objectives to observable indicators; the after-action architecture links findings to remediation and re-testing; the ENISA-based lifecycle provides a common process structure; and the regulatory pathway connects external requirements with national capability validation.

The evidence therefore supports the conclusion that a structured national programme provides a more suitable basis for resilience-oriented exercise governance than isolated activity because it creates explicit ownership, common information structures, repeatable lifecycle logic, longitudinal visibility, and formal mechanisms for corrective action and verification. In this sense, H2 is **supported within the scope and methodological boundaries of the dissertation**.

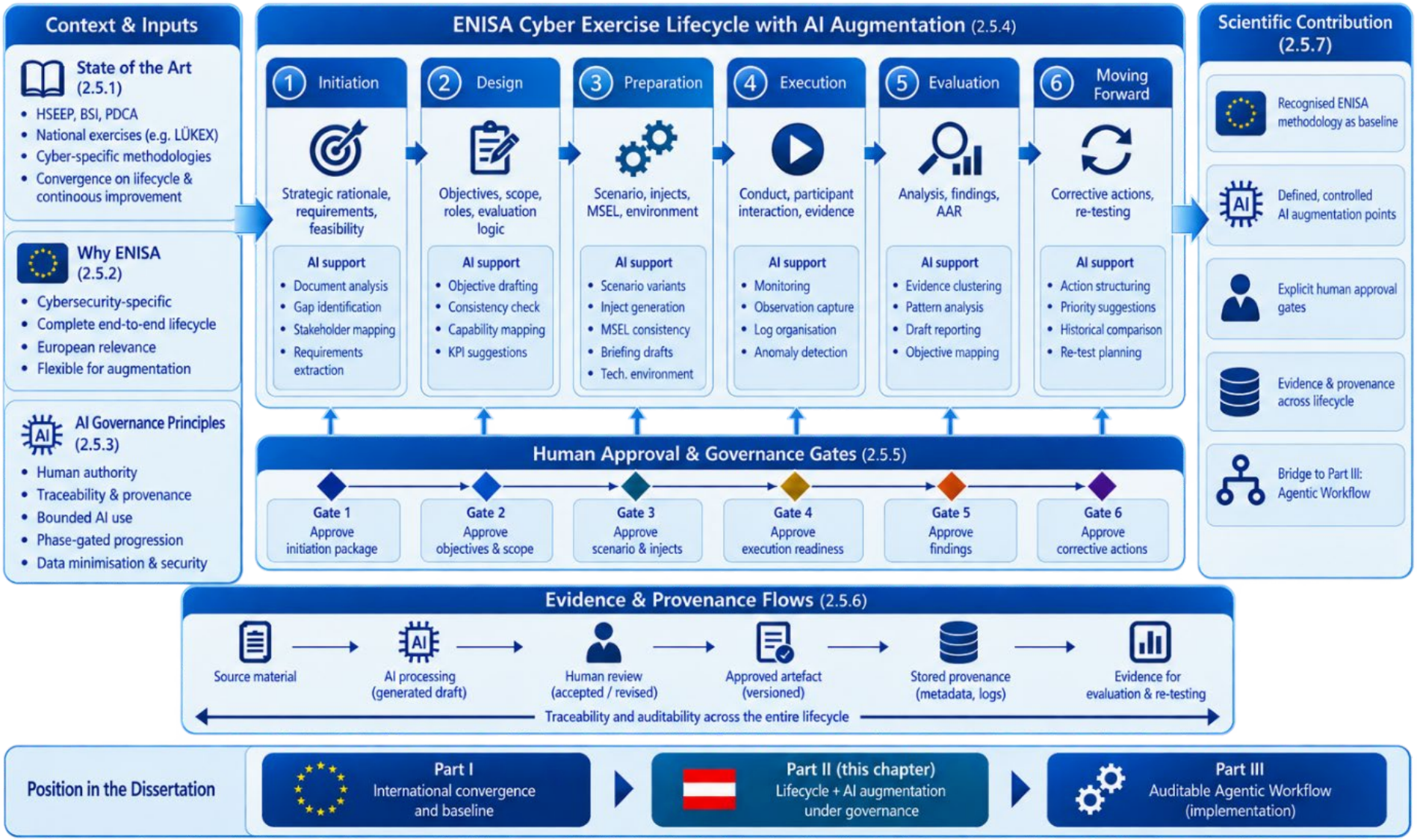


FIGURE 24 - ENISA-BASED AI-ENHANCED CX LIFECYCLE - CHAPTER OVERVIEW

This conclusion should nevertheless be interpreted cautiously. The dissertation does not claim that the proposed governance model has already produced measurable national cyber resilience improvement across Austria, nor that Austria represents a universally transferable national model. Austria is used as a reference environment in which the interaction between European requirements, national governance, sectoral structures, and exercise practice can be examined. The proposed model is constructed from this case in combination with the international and European findings of Part I, established exercise methodologies, and the regulatory and policy analysis. Its scientific validity therefore lies primarily in its coherence, traceability, applicability, and ability to integrate heterogeneous exercise activity into a common capability-development architecture.

Similarly, the dissertation does not establish a causal relationship according to which adoption of the model would automatically increase resilience. National cyber resilience is influenced by numerous factors beyond exercising, including resources, institutional maturity, workforce capability, threat exposure, technical architecture, political priorities, and organizational culture. The contribution of Part II is consequently narrower and more defensible: it demonstrates that a structured governance architecture provides the necessary conditions for exercise activities to become more comparable, auditable, evidence-oriented, and connected to verified improvement.
Part II therefore answers MRQ2 by showing that a national defence-oriented cyber exercise ecosystem can be designed through the integration of governance roles, common classification and data structures, coordinated portfolio management, objective-linked measurement, lifecycle discipline, differentiated information governance, after-action remediation, and re-testing. The national model does not seek to centralise all exercise activity or impose one exercise format. Instead, it provides a federated framework within which diverse organizations and sectors can contribute to a common national capability-development process.

The scientific contribution of Part II is thus broader than the National Cyber Exercise Governance Model alone. The Austrian reference case provides the national contextual foundation; the Cyber Exercise Taxonomy and Minimum Dataset provide the common information architecture; the National Cyber Exercise Governance Model and Programme Blueprint provide the institutional and programme structure; the ENISA-based AI-Enhanced Cyber Exercise Lifecycle provides the repeatable methodological process; and Regulatory Pathway II explains how European requirements are translated into national governance and exercise-relevant capabilities. These components are mutually reinforcing and together constitute the national-layer response to the dissertation's central research problem.

Part II therefore establishes the middle layer of the dissertation's cumulative research architecture. Part I identified **what is converging** at international and European level. Part II has shown **how this emerging baseline can be translated into national governance, common information structures, measurable exercise processes, and continuous improvement mechanisms**. The remaining question concerns operational implementation: how can the AI augmentation defined at lifecycle level be translated into concrete organizational processes without weakening traceability, accountability, security, or human authority?

This question provides the transition to Part III. The national layer defines the governance environment and the permitted augmentation conditions; the organizational and operational layer must now determine how these conditions can be implemented in practice through an auditable, bounded, and human-controlled Agentic Workflow. In this way, the progression from Part II to Part III follows directly from the logic of H2 to H3: national governance establishes the control

framework, while organizational operationalisation determines how AI support can function within that framework.

3 Part III – Organizational and Operational Research Layer

3.1 Introduction: From Lifecycle-Level AI Support to Organizational Operationalisation

Part III operationalises the AI augmentation principles developed at national lifecycle level. Part II defined where AI may support the ENISA exercise process and under which governance conditions. Part III addresses the implementation question: how can an organization translate those permitted augmentation points into a reproducible, auditable, and human-controlled workflow?
The organizational problem is not simply that exercise planning is time-consuming. Unstructured use of generative AI can produce inconsistent artefacts, introduce undocumented assumptions, weaken source grounding, and obscure responsibility for final decisions. These risks are particularly important in regulated and defence-oriented environments, where exercise artefacts may contain sensitive information and may later be used to justify findings, remediation, or assurance claims.
The proposed solution is an Auditable Agentic Workflow. Instead of treating AI as a general-purpose assistant, the workflow decomposes exercise planning and evaluation into bounded responsibilities with explicit inputs, expected outputs, policy constraints, provenance requirements, and human approval gates. The workflow operationalises the lifecycle principle 'AI recommends; humans decide'.
Part III contributes directly to H3 and MRQ3 by defining the architecture, control mechanisms, implementation logic, validation criteria, and organizational regulatory pathway required for responsible AI-supported cyber exercise production.



FIGURE 25 - OVERVIEW OF ORGANIZATIONAL RESEARCH LAYER

3.2 Auditable Agentic Workflow (C3)

The relevance of generative artificial intelligence to military and national-security decision support has also been examined in Hungarian defence research. Nagy (2025) investigates its applications in national-security analysis and assessment, with particular emphasis on supporting professional decision-making. Building on this research context, the present dissertation extends the application of generative AI towards defence-oriented cyber exercise preparation through an implemented and

empirically evaluated, human-supervised agentic workflow that integrates retrieval-augmented generation, specialised agents, and auditable exercise artefacts.[107]

The AI-enhanced lifecycle defines where AI can support the ENISA exercise process and under which governance conditions this support should operate. It does not yet prescribe a specific technical implementation. The subsequent Agentic Workflow translates these lifecycle-level augmentation points into specialized, phase-gated AI components with defined inputs, outputs, and human approval mechanisms.

Prior work involving LLM-supported compliance reasoning demonstrated the potential of combining generative AI with structured knowledge representations and explicit compliance constraints rather than relying solely on unrestricted natural-language generation (Turaga et al., 2025[108]). The workflow proposed here applies a related governance principle to cyber-exercise engineering: AI components operate within bounded tasks, structured inputs and outputs, provenance requirements and mandatory human approval points.

Concept and Rationale for Agentic Workflow Design

The term *agentic workflow* describes an approach to applying artificial intelligence in which work is decomposed into a set of bounded, goal-directed tasks that are executed by specialized components ("agents") in a controlled sequence. In contrast to single-prompt text generation, an agentic workflow is designed as a *process*: it specifies what needs to happen, in what order, under which conditions, and with which outputs. Each agent is assigned a clearly delimited responsibility (e.g., defining scope, drafting objectives, engineering injects for building a cyber exercises), receives explicit inputs, and produces outputs in a structured, reviewable form. Crucially, agency in this context does not imply autonomous decision-making outside human oversight. It denotes that the system can take initiative within constraints, for example by iterating over checklists, reconciling inconsistencies, or proposing artefact drafts, while humans remain responsible for approving, rejecting, or refining the results.

Agentic workflows have gained relevance because many applied AI use cases fail not due to model capability alone, but due to weak process integration. When an AI system is used as a general-purpose assistant, results often remain difficult to reproduce: outputs vary across prompts, assumptions are implicit, and intermediate steps are lost. This is a known mismatch for domains where work products must be traceable, defensible, and consistent across cycles. **Agentic workflows address this mismatch by embedding AI into an explicit pipeline: inputs are versioned, outputs are standardised, and quality gates are defined**. In practice, the workflow behaves less like an "assistant" and more like a structured planning assembly line where AI contributes drafts, checks, and transformations at defined points, and humans retain decision authority at each gate.

A defining characteristic of agentic workflows is *phase-gating*. Instead of letting the system generate an entire plan in one step, the workflow moves through phases that mirror the structure of a real planning process. Each phase produces artefacts that are necessary inputs to the next phase, and the transition is controlled by explicit "go/no-go" criteria. These criteria can include completeness requirements (e.g., all roles in scope have assigned responsibilities), consistency checks (e.g., each objective is linked to at least one inject and at least one evaluation measure),

[107] Nagy, V. (2025). A generatív mesterséges intelligencia felhasználásának területei a nemzetbiztonsági elemző-értékelő munkában, különös tekintettel a döntéstámogatásra. In T. Kenedli & S. Magyar (Eds.), *A mesterséges intelligencia hatásainak specifikus vizsgálata* (pp. 348–369). Katonai Nemzetbiztonsági Szolgálat. https://knbsz.gov.hu/kiadvanyok

[108] Turaga, V. S. P., Gopali, S., Kieseberg, P., Páhi, T., Tjoa, S., Jones, K. S., & Siami Namin, A. (2026). ACT: A conceptual framework for understanding and controlling trustworthiness in controllable AI. IEEE Access, 14, 84016–84066.

and compliance constraints (e.g., data handling limitations are respected, restricted information is excluded, and collection points are defined). If a gate fails, the workflow does not continue by default; instead, it either requests human revision or triggers a bounded retry with documented changes. This design principle is central for ensuring that AI support strengthens, rather than undermines, accountability.

In cybersecurity exercise planning, the relevance of agentic workflows follows directly from the nature of the planning problem. Cyber exercises are increasingly expected to be repeatable, comparable, and auditable ,particularly when they support sectoral resilience targets and national programmes. At the same time, the preparation of a high-quality exercise remains labour-intensive. Planners must translate strategic intent into measurable objectives, define scope and roles, develop a plausible scenario, design injects and artefacts, and prepare an evaluation plan that produces actionable improvement measures. In practice, these tasks are often carried out under time pressure, with limited facilitator capacity, and with heterogeneous planning maturity across organizations. The result is a set of frictions that repeatedly appear across exercise programmes: artefacts differ in quality and granularity, documentation is uneven, and scaling exercises across multiple entities becomes difficult because comparability requires common structures, shared assumptions, and consistent traceability.

Agentic workflows are relevant in this context because the exercise planning process is inherently modular and sequential. Each planning step produces concrete deliverables that become prerequisites for later steps: scope informs stakeholder mapping; stakeholder mapping constrains objectives; objectives constrain scenario design; scenario design yields injects; injects define what must be observed; and observation design shapes evaluation and improvement actions. This makes exercise preparation well-suited to a controlled pipeline rather than an open-ended generation task. By mapping each planning step to a specialised agent, the workflow can support planners precisely where effort tends to be repetitive, template-driven, or dependent on systematic cross-checking. Examples include drafting first versions of scope statements and role matrices, translating scenario backbone elements into a Master Scenario Event List (MSEL), or enforcing consistent linkages between objectives, injects, evidence sources, and evaluation criteria. The intention is not to replace the judgement and contextual understanding of experienced planners, but to reduce friction in the parts of the process that are structurally stable and therefore amenable to standardisation.

Beyond efficiency gains, a key reason for adopting an agentic workflow is *auditability*. In regulated or safety-critical environments, it is not sufficient to claim that AI helped; it must be possible to explain *how* it helped, what information it used, and who approved the resulting decisions. Agentic workflows support this requirement by design. Each agent operates with explicit inputs and produces outputs that can be stored, compared, and reviewed. Assumptions can be required as first-class fields rather than hidden in prose. Decision points can be logged, including which outputs were accepted, which were revised, and why. This enables provenance tracking and supports the creation of a defensible evidence trail. Moreover, because the workflow is phase-gated, the artefacts can be validated incrementally. Completeness checks can be automated (e.g., ensuring required sections exist), and internal consistency can be verified computationally (e.g., ensuring each inject maps to at least one objective and to an intended observation). These properties are not merely operationally convenient; they are methodologically important when the workflow is used as a research instrument, because they enable repeatable experiments and measurable outcomes.

Within this dissertation, the agentic workflow is therefore treated as more than an implementation choice. It is used as a methodological mechanism to operationalise and test AI support along the exercise lifecycle. Each agent is defined through a bounded responsibility and a set of expected outputs that correspond to recognised exercise deliverables. This makes it possible to evaluate the contribution of AI at the level where exercise quality is typically judged: artefact completeness,

internal consistency, linkage quality between objectives and injects, and the practicality of evaluation outputs. It also enables controlled comparisons across cases, because the same workflow structure can be applied to different organizations or sectors while keeping the artefact schema stable. In other words, the workflow supports comparability not only of exercises, but also of the AI-assisted planning process itself.

The agentic workflow used in this research is implemented as a set of specialised agents orchestrated by a phase controller. The **Orchestrator** (Phase Controller) runs the lifecycle in the correct order, triggers each agent, enforces go/no-go gates, and manages bounded retries. The **Requirements & Compliance Agent** extracts constraints and requirements (such as legal and regulatory expectations, data handling limitations, and operational boundaries) and flags gaps and risks that must be addressed early to prevent invalid exercise designs. The **Scope & Stakeholder Agent** defines the exercise scope, including what is explicitly in and out, which assumptions apply, which dependencies exist, and which stakeholders, player roles, and facilitators are required. The **Objectives & KPIs Agent** translates strategic intent into measurable objectives and connects them to success criteria, indicators, and evidence sources, so that evaluation is grounded in observable outcomes rather than narrative plausibility alone. The **Scenario Architect Agent** builds the scenario backbone by defining the state of the world, the threat actor profile, storyline phases, and decision points, ensuring alignment with the objectives and the identified constraints. The **MSEL-Inject Engineer Agent** converts the storyline into a Master Scenario Event List, mapping each inject to objectives, target roles, expected actions, and the data that should be observed. Finally, the **Evaluation Agent** produces the evaluation design: it defines what evidence is collected, how observations are mapped to objectives and indicators, how findings are structured (e.g., after-action review outputs), and how improvement measures are formulated so that they can be tracked and re-tested in subsequent cycles.

By embedding AI into an auditable, human-supervised workflow aligned with a recognised exercise methodology, this dissertation aims to produce evidence on where AI can meaningfully improve efficiency and artefact quality in cyber exercise preparation, and where its use introduces risks or unacceptable trade-offs. **The central proposition is not that AI "writes" better exercises, but that a phase-gated, role-separated workflow can make planning more consistent, more scalable, and more defensible, while preserving methodological rigour, traceability, and the human responsibility that effective exercises require.**

These requirements distinguish the contribution from a sequence of independent prompts. The workflow is designed as a controlled production system in which the quality and provenance of each artefact affects whether the next stage is allowed to proceed.
The design requirements are: bounded responsibilities for each AI-supported component; explicit structured inputs; structured outputs corresponding to recognised exercise artefacts; source grounding and explicit assumptions; policy and data-access constraints; mandatory human approval for consequential artefacts; phase-gated progression; provenance and versioning; persistent decision logging; and the ability to trace downstream artefacts back to requirements and evidence.
The architecture is derived from the research problem rather than from a preference for a particular AI technology. The workflow must preserve the properties identified as necessary across Parts I and II: comparability, traceability, governance compatibility, bounded responsibility, evidence quality, and continuous improvement.

Design Requirements for the Auditable Agentic Workflow

3.2.1.1 *Proposed Agentic Structure for Cyber Exercises*

Cyber exercises are increasingly expected to be repeatable, comparable, and auditable ,especially when they support sectoral resilience targets and national programmes. At the same time, planning and preparing a high-quality exercise remains a labour-intensive activity. Exercise teams must translate strategic intent into measurable objectives, define scope and roles, develop a plausible scenario, design injects and artefacts, and prepare an evaluation

plan that produces actionable improvement measures. In practice, this work is often conducted under time pressure, with limited facilitator capacity and heterogeneous planning maturity across organizations. These conditions create well-known frictions: inconsistent artefact quality, uneven documentation, and difficulties in scaling exercises across multiple entities or sectors while maintaining a common standard.

This dissertation investigates whether an **agentic workflow** ,a structured, phase-gated orchestration of specialised AI “agents” under explicit human control ,can address these frictions without compromising accountability. The motivation is not to replace the judgement and contextual understanding of experienced exercise planners, but to support them where planning work is repetitive, template-driven, or dependent on systematic cross-checking. Examples include drafting first versions of scope statements and role matrices, transforming storyline elements into a Master

From Prompting to Process: Agentic Workflows for Cyber Exercises

An agentic workflow is a structured, phase-gated orchestration of specialised AI components that each perform a bounded planning task and produce defined, reviewable outputs, rather than generating an exercise plan in one open-ended step.
It is relevant because cyber exercises increasingly need repeatability, comparability, and auditability, while planning is time-critical and often suffers from inconsistent artefact quality and documentation. The approach is particularly suitable for cyber exercises because planning is modular and sequential (scope, objectives, scenario backbone, MSEL/injects, and evaluation artefacts depend on each other) so the process can be modelled as a controlled pipeline with explicit go/no-go checkpoints.

In this dissertation, the agentic workflow is used as a research instrument to operationalise AI support across lifecycle phases, with each agent producing structured deliverables (e.g., scope/roles, objectives/KPIs, scenario backbone, MSEL, evaluation plan) under human approval. This enables systematic validation of completeness, internal consistency (e.g., injects linked to objectives), and decision provenance through versioned logs, allowing measurable and reproducible evaluation of benefits and trade-offs.

Scenario Event List (MSEL), or ensuring that objectives, injects, and evaluation measures remain consistently linked. An agentic approach is particularly suitable for cybersecurity exercises because the planning process is inherently modular and sequential: each phase produces concrete artefacts that become inputs for subsequent phases. This structure allows the workflow to be modelled as a controlled pipeline with explicit “go/no-go” checkpoints, rather than as an open-ended text generation task.

Methodologically, the agentic workflow is used as a **research instrument** to operationalise and test AI support along the exercise lifecycle. Each agent is defined by a bounded responsibility, receives clearly specified inputs, and must produce structured outputs that correspond to recognised exercise deliverables (e.g., exercise plan, scenario backbone, MSEL, evaluation plan). This design enables systematic validation: completeness can be checked automatically, internal consistency can be verified (e.g., injects map to objectives), and the provenance of decisions can be documented through versioned logs. The result is a setup that is measurable and reproducible ,two requirements that are central to scientific evaluation and to the practical adoption of AI in regulated and safety-critical contexts.

By embedding AI into an auditable, human-supervised workflow aligned to a recognised methodology, this dissertation aims to contribute evidence on where AI can meaningfully improve

efficiency and artefact quality in cyber exercise preparation, and where its use introduces risks or unacceptable trade-offs. Ultimately, the agentic workflow approach is examined as a pathway to scale cyber exercise planning across organizations and sectors while preserving methodological rigour, traceability, and the human responsibility that effective exercises require.

The agentic workflow uses the following agents:

1. **Orchestrator / Phase Controller**
   Runs the lifecycle in the correct order, triggers each agent, enforces go/no-go gates, and manages retries.
2. **Requirements & Compliance Agent**
   Extracts constraints and requirements (legal/regulatory, data handling, operational limits) and flags gaps/risks.
3. **Scope & Stakeholder Agent**
   Defines exercise scope (in/out, assumptions, dependencies) and builds the stakeholder + player/role matrix.
4. **Objectives & KPIs Agent**
   Creates measurable exercise objectives and links them to success criteria, KPIs/indicators, and evidence sources.
5. **Scenario Architect Agent**
   Builds the scenario backbone: state-of-the-world, threat actor profile, storyline phases, and decision points aligned to objectives and constraints.
6. **MSEL / Inject Engineer Agent**
   Converts the storyline into a Master Scenario Event List (inject timeline), mapping each inject to objectives, target roles, expected actions, and data to observe.
7. **Evaluation Agent**
   Produces the evaluation plan: objective → indicator → metric mapping, data collection instruments/roles, analysis approach, and reporting outputs (AAR-ready).

Cyber exercises are increasingly expected to be repeatable, comparable, and auditable ,especially when they support sectoral resilience targets and national programmes. At the same time, planning and preparing a high-quality exercise remains a labour-intensive activity. Exercise teams must translate strategic intent into measurable objectives, define scope and roles, develop a plausible scenario, design injects and artefacts, and prepare an evaluation plan that produces actionable improvement measures. In practice, this work is often conducted under time pressure, with limited facilitator capacity and heterogeneous planning maturity across organizations. These conditions create well-known frictions: inconsistent artefact quality, uneven documentation, and difficulties in scaling exercises across multiple entities or sectors while maintaining a common standard.

This dissertation investigates whether an **agentic workflow**, a structured, phase-gated orchestration of specialised AI “agents” under explicit human control, can address these frictions without compromising accountability. The motivation is not to replace the judgement and contextual understanding of experienced exercise planners, but to support them where planning work is repetitive, template-driven, or dependent on systematic cross-checking. Examples include drafting first versions of scope statements and role matrices, transforming storyline elements into a Master Scenario Event List (MSEL), or ensuring that objectives, injects, and evaluation measures remain consistently linked. An agentic approach is particularly suitable for cybersecurity exercises because the planning process is inherently modular and sequential: each phase produces concrete artefacts that become inputs for subsequent phases. This structure allows the workflow to be modelled as a controlled pipeline with explicit “go/no-go” checkpoints, rather than as an open-ended text generation task.

Methodologically, the agentic workflow is used as a **research instrument** to operationalise and test AI support along the exercise lifecycle. Each agent is defined by a bounded responsibility,

receives clearly specified inputs, and must produce structured outputs that correspond to recognised exercise deliverables (e.g., exercise plan, scenario backbone, MSEL, evaluation plan). This design enables systematic validation: completeness can be checked automatically, internal consistency can be verified (e.g., injects map to objectives), and the provenance of decisions can be documented through versioned logs. The result is a setup that is measurable and reproducible ,two requirements that are central to scientific evaluation and to the practical adoption of AI in regulated and safety-critical contexts.

By embedding AI into an auditable, human-supervised workflow aligned to a recognised methodology, this dissertation aims to contribute evidence on where AI can meaningfully improve efficiency and artefact quality in cyber exercise preparation, and where its use introduces risks or unacceptable trade-offs. Ultimately, the agentic workflow approach is examined as a pathway to scale cyber exercise planning across organizations and sectors while preserving methodological rigour, traceability, and the human responsibility that effective exercises require.

3.2.1.2 *Initiation with AI layer*

The Initiation phase establishes the foundation of a cybersecurity exercise by clarifying *why* the exercise is conducted, what type of exercise is appropriate, who must be involved, and whether the planned activity is feasible under real-world constraints. This phase is concerned with early alignment (strategic intent and expected value), resource and stakeholder readiness, and the establishment of planning baselines that prevent downstream rework.

Phase 1 objectives and core planning activities

In practice, Initiation consolidates three categories of decisions:

1. **Intent and exercise framing.** The organising body defines the purpose of the exercise (e.g., capability validation, coordination rehearsal, decision-making under stress) and selects an appropriate exercise type (discussion-based, operation-based, or hybrid). This framing determines the expected level of realism, the required infrastructure, and the level of organizational disruption that can be tolerated.

2. **Stakeholder and role identification.** The phase identifies the exercise owner/sponsor, the planning team, and the participating stakeholder groups. This includes an initial view of key roles (e.g., decision-makers, incident response leadership, technical responders, legal/communications) and of the functions that must be represented for the exercise to meaningfully address the stated purpose.

3. **Constraints and feasibility assessment.** Initiation captures legal, operational, and organizational constraints that shape what can be safely and lawfully exercised (e.g., data protection boundaries, restrictions on using production systems, confidentiality constraints, availability of personnel and time windows). The feasibility check aims to ensure that scope and ambition remain consistent with resources, timelines, and organizational readiness.

The main output of Initiation is an initial, structured planning baseline that can be iteratively refined in later phases, most notably a first version of the exercise plan (or exercise concept document), together with clearly stated assumptions, boundaries, and known risks.

Agentic workflow implementation in Phase 1

To operationalise Initiation in a reproducible way, this dissertation models the phase as a **phase-gated agentic workflow**. The workflow does not treat “planning” as a single text generation task. Instead, it decomposes Initiation into bounded responsibilities executed by specialised agents that produce structured artefacts. The approach mirrors established planning practice: each step yields a concrete intermediate product that becomes input for the next step, while a go/no-go checkpoint enforces completeness before proceeding.

In Phase 1, the workflow uses four functional roles:

1. **Orchestrator (Phase Controller).** The orchestrator initiates the process, enforces the Initiation sequence, and ensures that all required inputs are collected in a consistent form (sector context, exercise owner, high-level intent, time constraints). It also manages iterative refinement: if a later step detects missing information, the orchestrator routes the workflow back to the responsible agent rather than continuing with incomplete artefacts.
2. **Requirements & Compliance Agent.** This agent extracts and formalises constraints that govern the exercise design, such as data handling requirements, regulatory considerations, and operational safety limits. Its output is a constraint set that can be re-used as validation rules in later phases (e.g., preventing scenario elements or injects that would require prohibited datasets or unsafe actions). Early formalisation is essential because these constraints often drive feasibility and scope.
3. **Scope & Stakeholder Agent.** This agent produces an initial scope statement (in-scope/out-of-scope, assumptions, dependencies) and identifies key stakeholder groups and roles needed to meet the exercise purpose. The output is deliberately “first-pass”: it is sufficient to establish boundaries and participation expectations, while leaving detailed role descriptions and exercise control structures to subsequent phases.
4. **Feasibility synthesis (within the orchestrated step).** Using the constraints and scope outputs, the workflow produces a feasibility summary: resource needs, anticipated bottlenecks, and risks that must be mitigated before detailed design begins. This synthesis is treated as a decision-support artefact rather than a final judgement; responsibility for acceptance remains with the human exercise owner.

Go/No-Go gate and traceability

A central methodological element is the **Initiation go/no-go gate**, which functions as a quality control checkpoint. The gate evaluates whether the minimum viable planning baseline exists and is internally coherent. At minimum, the gate verifies that: (i) purpose and exercise type are defined, (ii) scope boundaries (in/out) are explicit, (iii) relevant stakeholders and participant groups are identified, (iv) constraints are documented (data-handling and operational), and (v) feasibility status and key risks are recorded.

If the gate fails, the workflow does not proceed to Design; instead, it loops back to the responsible agent(s) to repair the missing or inconsistent elements. If the gate passes, the workflow produces an Initiation package (Exercise Plan v0.1 and associated constraint/scope artefacts) and logs the decision to proceed.

To maintain auditability, the workflow captures **provenance information** for each agent step: timestamps, input versions, produced artefacts, and the human approval decision at the gate. This logging mechanism is not a secondary implementation detail; it is a research requirement that enables later comparison between baseline planning and AI-augmented planning, and it supports accountability in environments where exercise documentation may be reviewed by internal governance or external stakeholders.

Phase 1 outcome for subsequent phases

The outcome of Initiation is therefore not only a set of draft documents but a structured planning state that can be reused and validated in later phases. Design builds upon this state by turning the purpose into measurable objectives and a refined participant model, while Preparation uses the constraints and scope to ensure scenario realism, inject integrity, and evaluation alignment. By restricting Phase 1 to bounded responsibilities and enforcing a go/no-go gate, the agentic workflow aims to reduce avoidable rework, increase planning consistency, and provide a defensible baseline for empirical evaluation of time and quality effects in subsequent phases.

1. **Phase 1: Initiation**

Goal (ENISA): Decide *why* you run an exercise, choose the exercise type, identify resources and stakeholders, understand posture, and check feasibility. The main deliverable is the **Exercise Plan (first version)**.

2. **What happens in plain words**

You are setting the foundation: “What is this exercise about, who is involved, what are the boundaries, and can we realistically do it?”

3. **How the agents work here**

4. **Orchestrator**

- Starts the workflow, collects the basic input (sector, target audience, preferred date, exercise type idea).
- Calls the agents in the right order.
- Stops you at a **go/no-go checkpoint** to confirm you want to proceed. (ENISA recommends go/no-go points between phases.)

2. **Requirements & Compliance Agent**

- Asks: “What rules must we respect?”
- Produces a clear list like:
    - data protection boundaries (e.g., no real customer data),
    - regulatory focus (e.g., incident reporting obligations),
    - operational constraints (e.g., blackout windows, systems that must not be touched).
- Flags risks early (e.g., “this scenario would require data we cannot use”).

3. **Scope & Stakeholder Agent**

- Asks: “Who and what is included?”
- Produces:
    - in-scope vs out-of-scope,
    - key stakeholder list (sponsor, planning team, players, observers),
    - a first **player/team picture** (e.g., incident response, legal, comms), aligned with ENISA’s emphasis on early stakeholder mapping.

4. **Orchestrator (again)**

- Compiles these into the **Exercise Plan v0.1** (a first structured blueprint).
- Runs **Gate 1 (Initiation go/no-go)**:
    - Do we have a clear “why” (purpose)?
    - Do we know the exercise type (discussion-based / operation-based)?
    - Are scope + constraints written down?
    - Are stakeholders identified?
    - Is it feasible with our resources?

Human step: You review and approve the Gate 1 output.

3.2.1.3 *Design with AI layer*

The Design phase transforms the initial exercise concept developed during Initiation into a coherent and measurable exercise design. While Initiation clarifies *why* an exercise should be conducted and whether it is feasible, Design specifies *what exactly will be exercised*, *who will participate*, and *how success will be judged*. The quality of this phase largely determines whether later scenario and inject development will remain focused, comparable across organizations, and suitable for systematic evaluation.

Phase 2 objectives and core planning activities

Design consolidates the exercise intent into a set of design decisions that create a stable basis for Preparation:

1. **Definition of objectives and success criteria.** The central task of Design is to formulate exercise objectives that are specific enough to guide scenario development and evaluation. Objectives must be measurable and actionable; they should define what capability, coordination mechanism, or decision-making process is being tested. For each objective, the planning team establishes a preliminary view of what "success" looks like and which observable indicators could demonstrate achievement.
2. **Refinement of scope and participation model.** Design turns the initial scope statement into a refined boundary definition. This includes explicitly stating assumptions and dependencies, identifying interfaces between organizational units, and clarifying exclusions to avoid uncontrolled expansion of the exercise. In parallel, the participation model is formalised: key teams, roles, and decision levels are specified so that the exercise realistically reflects organizational structures and accountability lines.
3. **Establishment of evaluation foundations.** A recurring weakness in exercise programmes is late or ad-hoc evaluation planning. Design addresses this by defining an evaluation skeleton that links objectives to intended data sources and collection mechanisms. At this stage, evaluation does not require detailed instruments; rather, it establishes the logic that will later be operationalized during Preparation, ensuring that injects and expected actions are explicitly measurable.
4. **Initial communication and coordination planning.** Depending on exercise type and stakeholder environment, Design also initiates communication planning: internal communication channels, stakeholder briefing needs, confidentiality expectations, and coordination procedures for controllers, facilitators, and observers. These elements are critical for managing realistic information flows during the exercise without creating confusion about what is simulated and what is operational.

The primary outcome of Design is an updated and more detailed exercise plan, complemented by a structured set of objectives and a credible evaluation foundation. Together, these outputs enable scenario and MSEL development in the next phase without ambiguity regarding scope, roles, or measurement.

Agentic workflow implementation in Phase 2

In the proposed agentic workflow, the Design phase is implemented as a phase-gated sequence of specialised agents that refine the planning baseline created in Initiation. The workflow is designed to keep responsibilities bounded and outputs structured, allowing systematic validation and reducing the likelihood of inconsistencies that would otherwise propagate into Preparation.
The Design phase uses the following agents:

1. **Orchestrator (Phase Controller).** The orchestrator initiates the Design phase by loading the Phase 1 outputs (constraints, scope, stakeholder picture, feasibility notes) and executing the required Design steps in a controlled order. It coordinates iterative refinement when a design element is incomplete or inconsistent and enforces the Design go/no-go gate before the workflow proceeds to scenario development.

2. **Objectives & KPIs Agent.** This agent converts the exercise purpose into a set of measurable objectives, each paired with success criteria and candidate indicators. The agent produces objectives in a structured form, ensuring consistent wording and traceable identifiers. Importantly, the output is designed to be reusable: objective identifiers become tags that link later scenario phases, injects, and evaluation measures.
3. **Scope & Stakeholder Agent (refinement pass).** In Design, the scope agent revisits the Phase 1 scope statement and refines it using the objective set as a stabilising anchor. The agent clarifies assumptions, dependencies, exclusions, and interfaces between participating units, and it confirms that the participation model aligns with the objectives (e.g., if an objective involves executive decision-making, executive roles must be included). This refinement reduces “scope drift” in later phases.
4. **Evaluation Agent (evaluation skeleton).** The evaluation agent produces an early evaluation structure that links each objective to potential evidence sources and data collection approaches. At this stage, the deliverable is intentionally lightweight: it ensures that every objective is evaluable and that the planning team has a clear understanding of what will need to be observed or captured later. This step also provides a basis for validating the suitability of proposed objectives: if an objective cannot be measured or observed in the planned exercise type, it is revised before Preparation begins.

Depending on organizational context, the workflow may additionally include an early communication planning step; however, this can also remain part of the exercise plan compilation if a separate communication plan is not within the scope of the initial pilot.

Go/No-Go gate and traceability
The Design phase concludes with a **Design go/no-go gate** that validates whether the exercise is sufficiently specified to begin scenario and inject construction. The gate verifies that: (i) objectives exist, are measurable, and include success criteria; (ii) scope boundaries are explicit, including assumptions, dependencies, and exclusions; (iii) the participation model is defined and aligned to objectives; and (iv) an evaluation skeleton exists that links objectives to data sources or observation points.
If the gate fails, the workflow routes back to the responsible agent(s) ,for example, objectives are refined if they are not measurable, scope is tightened if boundaries remain unclear, or the evaluation skeleton is extended if an objective is not observable. If the gate passes, the workflow produces a Design package consisting of an updated exercise plan and structured objective and evaluation foundations, and it records a traceable approval decision.
Throughout the phase, provenance information is captured for each agent output, including timestamps, version identifiers, and change summaries. This traceability serves two purposes: it supports accountability in regulated environments and it enables empirical comparison between baseline and AI-augmented planning in subsequent evaluation of time, quality, and consistency outcomes.

Phase 2 outcome for subsequent phases
By the end of Design, the exercise has a stable definition of objectives, scope, roles, and evaluation logic. This enables Preparation to focus on the constructive work of scenario development and MSEL creation rather than repeatedly renegotiating boundaries or success criteria. In the agentic workflow, objective identifiers and scope constraints become control parameters that guide scenario generation, ensure that injects test intended capabilities, and prevent later inconsistencies between the narrative, the planned participant actions, and the evaluation plan.

3.2.1.4 *Preparation with AI layer*

The Preparation phase converts an exercise design into an executable package. While the Design phase defines objectives, scope, roles, and an evaluation logic, Preparation is the phase in which the exercise becomes operational: a coherent scenario is constructed, participant-facing stimuli are planned, and evaluation is translated from an abstract concept into concrete data collection and

analysis procedures. The phase is therefore decisive for training realism and for the credibility of subsequent exercise results. Weaknesses introduced here ,such as unrealistic pacing, injects that do not test the intended objectives, or an evaluation plan that cannot be executed ,directly degrade the usefulness and comparability of the exercise.

In most cases, synthetic media is generated for gaming, to improve services or to improve the quality of life, but the increase in synthetic media and improved technology has given rise to disinformation possibilities, including deepfakes.[109]

Phase 3 objectives and core planning activities

Preparation consolidates the design baseline into a runnable package through four main activity clusters:

1. **Scenario backbone development.** A plausible scenario is created that fits the sector context, scope boundaries, and participant model. The scenario backbone typically includes (i) a “state of the world” that describes the operational situation at exercise start, (ii) a threat actor and high-level attack narrative (or, in non-adversarial scenarios, a disruptive event narrative), and (iii) a storyline that unfolds through distinct phases and is explicitly aligned to the exercise objectives. Importantly, the scenario must respect constraints defined earlier (e.g., no real personal data, no unsafe actions, no undue reputational exposure) while still achieving sufficient realism to elicit meaningful decisions.

2. **Transformation of storyline into injects and timeline control.** The narrative is operationalized as a structured timeline of events and injects. This activity produces a master scenario event list (MSEL) that specifies what information is delivered, to whom, when, and through which channel. Injects serve as controlled stimuli that prompt participant actions and decisions; therefore, each inject must be associated with one or more objectives and should have a clear expectation regarding the decision or action it is intended to provoke. Preparation also determines inject pacing and escalation logic to avoid both overload and under-stimulation, ensuring that the exercise remains manageable and aligned to the intended learning and evaluation goals.

3. **Operational evaluation planning.** Preparation turns the evaluation skeleton created in the Design phase into a detailed evaluation plan. This includes specifying indicators and metrics for each objective, defining the exact data sources and collection methods (e.g., observer note templates, chat/ticket logs, time stamps, surveys), and assigning responsibilities for data capture and analysis. The evaluation plan must be feasible within the exercise format: it should not require evidence that cannot realistically be collected, and it must be robust enough to support defensible findings and improvement recommendations.

4. **Readiness, consistency, and stakeholder preparation.** Finally, Preparation ensures that all elements of the exercise package are consistent and implementable. This includes checking that scenario elements and injects do not violate constraints; that targeted roles exist and are briefed; that evaluation instruments match the planned events; and that the overall package can be executed as intended. Depending on exercise format, this may also involve briefings, dry runs, and final alignment between controllers, facilitators, and observers.

[109]Europol. https://www.europol.europa.eu/cms/sites/default/files/documents/Europol_Innovation_Lab_Facing_Reality_Law_Enforcement_And_The_Challenge_Of_Deepfakes.pdf

The outcome of Preparation is a complete exercise package that can be delivered during Execution without improvisation on core design elements, and that can generate measurable evidence for Evaluation and follow-up improvement.

Agentic workflow implementation in Phase 3

In the agentic workflow, Preparation is modelled as a structured sequence of specialised agents that progressively translate design intent into executable artefacts. The workflow is designed to reduce common quality issues by enforcing explicit linkages between objectives, storyline phases, injects, and evaluation measures.

The phase employs the following agents:

1. **Orchestrator (Phase Controller).** The orchestrator initiates Preparation by loading the validated outputs from Initiation and Design (constraints, refined scope, player model, objectives, and evaluation skeleton). It then calls the preparation agents in a fixed order and manages iterations if consistency checks fail. The orchestrator also enforces the Preparation go/no-go gate before the package is accepted as runnable.

2. **Scenario Architect Agent.** This agent constructs the scenario backbone. It produces a structured narrative consisting of the state of the world, a threat actor profile (or disruptive event profile), and storyline phases. The agent's output is explicitly conditioned on scope and constraints: it must remain within in-scope systems and actors, avoid prohibited data use, and respect operational limits. In addition, the scenario is anchored to the objective set by associating storyline phases with intended decisions and objective identifiers, creating traceable links that can later be used for evaluation mapping.

3. **MSEL / Inject Engineer Agent.** This agent operationalises the storyline as a time-based MSEL. It decomposes the narrative into injects that specify delivery channel, target role, inject content, expected action/decision, and objective linkage. The MSEL agent also produces fallback injects and controller notes where appropriate to maintain control over exercise flow. By tagging injects with objective identifiers, the workflow makes it possible to verify that each objective is exercised through concrete stimuli rather than being a purely declarative statement.

4. **Evaluation Agent (full plan).** The evaluation agent converts the design-stage evaluation skeleton into a full evaluation plan. For each objective, it defines indicators and metrics, specifies data sources and collection instruments, and assigns roles responsible for data capture. A critical task in this step is alignment with the MSEL: evaluation measures are linked to relevant inject identifiers so that observers know when and where evidence should be collected. The evaluation plan also includes a basic analysis approach and reporting outputs, ensuring that the collected data can be transformed into actionable findings.

5. **Requirements & Compliance Agent (constraint re-check).** In Preparation, the requirements agent plays a verifying role. It re-checks scenario elements and inject content against the previously defined constraints and flags violations. This step is essential to prevent realistic scenario writing from inadvertently crossing legal or operational boundaries, and it supports the thesis requirement of maintaining human accountability and auditability in an AI-supported workflow.

6. **Consistency & Risk Checks Agent (or validation node).** The workflow includes an explicit consistency step ,implemented as an agent or, preferably, as a deterministic validation node ,that checks internal coherence across artefacts. Typical checks include: every inject targets an existing role; every inject references at least one valid objective; pacing is plausible for the exercise duration; evaluation measures reference existing objectives and (where used) inject identifiers; and no inject conflicts with constraints. This

step turns quality expectations into verifiable rules and reduces the risk of silent inconsistencies.

Go/No-Go gate and traceability

Preparation concludes with a **Preparation go/no-go gate**. The gate verifies that the exercise package is complete and internally consistent, and that it can be executed without relying on ad-hoc improvisation. At minimum, the gate requires: (i) an approved scenario backbone with explicit linkage to objectives; (ii) a populated MSEL with injects mapped to objectives and roles; (iii) an evaluation plan with clear instruments, data sources, and responsible roles; and (iv) confirmation that constraints are respected throughout the package.

If the gate fails, the workflow routes back to the agent responsible for the failing component (scenario, MSEL, or evaluation) and repeats the generation and validation cycle until the package meets the acceptance criteria. If the gate passes, the artefacts are versioned as the final Preparation outputs and are ready for execution.

Throughout the phase, the workflow maintains a provenance record of each agent step, including version identifiers, timestamps, and change summaries. This traceability enables systematic evaluation of the workflow itself (e.g., where time is saved, where iteration occurs, and which artefacts benefit most from agentic support) and provides an auditable planning history for contexts where exercise documentation must withstand internal or external review.

Phase 3 outcome and boundary to later phases

At the end of Preparation, the exercise has transitioned from a conceptual design to a runnable package: the narrative, inject timeline, and evaluation instruments are aligned and constrained by agreed boundaries. This defines a clean boundary to later lifecycle phases. Execution can focus on facilitation and realistic interaction, while Evaluation and Moving Forward can rely on structured evidence produced according to the evaluation plan. In the scope of this dissertation, Preparation is therefore the final phase in which the agentic workflow is directly tested: it is the point where planning quality becomes observable in the completeness, coherence, and measurability of the produced artefacts.

1. **Phase 3: Preparation**

Goal (ENISA): Build a realistic scenario, translate it into a controlled timeline (MSEL), define the evaluation strategy (indicators/metrics/data collection), and prepare stakeholders. At the end, **Exercise Plan + Evaluation Plan + MSEL should be ready**.

2. **What happens in plain words**

Now you write the “story” of the incident and prepare exactly what participants will receive and when ,plus how you will measure performance.

3. **How the agents work here**

4. **Scenario Architect Agent**

- Uses ENISA’s scenario elements:
    - **threat actor**, **state of the world** (what happened before), and **storyline** (what happens during).

- Ensures realism and avoids overload (ENISA warns against being too dramatic or confusing).

- Outputs the **Scenario Backbone**.

2. **MSEL / Inject Engineer Agent**

- Transforms the storyline into ENISA's hierarchy:
    - **events → incidents → injects**
- Produces the **MSEL**: a timeline that says:
    - what inject is delivered,
    - through which channel (email, call, SIEM alert…),
    - to which role,
    - expected reaction/decision,
    - which objective it tests.

(ENISA calls the MSEL the "conductor's score".)

3. **Evaluation Agent (full plan)**

- Converts objectives into:
    - indicators (what you measure),
    - metrics (how you measure),
    - and **data collection tools** (where data comes from).

- Links evaluation to inject expected actions (ENISA explicitly recommends connecting the MSEL with the evaluation strategy).

4. **Requirements & Compliance Agent (quick compliance re-check)**

- Verifies the scenario and injects don't violate constraints:
    - no real-world data leakage,
    - no forbidden "live malware on production," etc.
- Flags anything that breaks your rules.

5. **Orchestrator → Gate 3 (Preparation go/no-go)**
Checks:

- Scenario exists and matches objectives and scope,
- MSEL exists and injects map to objectives,
- Evaluation plan exists with roles/tools and analysis approach,
- Consistency is good (no inject contradicts constraints; target roles exist; pacing is reasonable).

Human step: Final approval. If you reject, the Orchestrator routes back to the responsible agent (scenario vs MSEL vs evaluation).

3.2.1.5 *Evaluation*

To validate AI support in cyber exercise planning credibly - particularly in defence-oriented settings - evaluation must go beyond subjective usability impressions and focus on measurable outputs and traceable decision processes. A recent systematic literature review on AI in military cyber security highlights that a central challenge in this domain is not only model capability, but the *rigour of evaluation*: studies often rely on standard performance indicators (e.g., confusion-matrix-derived metrics) while simultaneously acknowledging the need for evaluation approaches that reflect operational constraints, robustness, and mission relevance. The paper "The Role of AI in Military Cyber Security: Data Insights and Evaluation Methods" [110] surveys how AI is being applied in military cyber security, with a specific focus on datasets and sources and, evaluation methods used to assess AI-based cyber security solutions in military contexts.

The review further underlines a recurring structural limitation in defence contexts: access to realistic datasets is restricted, and as a consequence, evaluation frequently depends on synthetic or simulated data and on controlled environments.

This observation is directly relevant for AI-supported cyber exercise planning because the artefacts produced during planning (scope definitions, objective sets, MSELs, evaluation plans) represent a form of "operational data" whose quality and traceability determine whether AI assistance is acceptable in regulated or safety-critical contexts.

In response, this dissertation adopts an evaluation design that treats the agentic workflow as a measurable intervention on the exercise lifecycle rather than as an unstructured assistant. The workflow outputs are assessed using **artefact-based and process-based criteria** that align with the evaluation discipline advocated in military AI cyber security research: (i) *output quality* that can be verified against defined requirements, and (ii) *operational suitability* measured through feasibility and accountability properties.

Concretely, artefact quality is evaluated through completeness and structure checks (e.g., whether required fields and deliverables are present), as well as internal consistency tests (e.g., whether each inject is mapped to at least one objective and associated evidence source). These checks operationalise quality in a way that supports repeatability across cases and reduces the reliance on purely narrative assessment. In parallel, the evaluation includes workload-oriented measures, such as planning time, number of revision cycles, and facilitator effort indicators, to ensure that any quality gains do not come at the cost of unacceptable operational burden.[111]

A defining requirement for defence-oriented applicability is **traceability**. The literature on military AI cyber security repeatedly emphasises that evaluation must account for robustness and operational trustworthiness, not merely "accuracy" in a narrow sense.

Accordingly, this dissertation evaluates traceability through the presence and quality of provenance records: versioned artefacts, logged human approvals at phase gates, and documented rationale for major changes (including AI-generated suggestions that were accepted or rejected). This creates an audit-ready chain that links strategic intent to produced deliverables and to evaluation outcomes. Taken together, the selected metrics and evidence design implement a controlled validation approach consistent with the broader direction in defence-relevant AI cyber security research: measurable, reproducible assessment grounded in operational constraints, rather than informal perceptions of usefulness

[110] Maathuis, C., & Cools, K. (2025). The Role of AI in Military Cyber Security: Data Insights and Evaluation Methods. Procedia Computer Science, 254, 191-200.

[111] Maathuis, C., & Cools, K. (2025). The Role of AI in Military Cyber Security: Data Insights and Evaluation Methods. *Procedia Computer Science*, *254*, 191-200.

3.2.1.6 *Summary: Merging the ENISA Cyber Security Exercise Methodology, the AI-Enhanced Lifecycle, and the Agentic Workflow*

The release of the ENISA Cyber Security Exercise Methodology provided an important opportunity to strengthen the methodological foundation of this dissertation. Rather than maintaining a separate lifecycle logic in parallel to a recognised European reference model, the present work adopts ENISA's end-to-end exercise methodology as the primary backbone for exercise planning, execution, evaluation, and improvement. At the same time, this dissertation does not stop at adopting ENISA as a descriptive best practice. Its contribution lies in extending the ENISA lifecycle in two ways: first, by introducing an explicit AI support layer across all phases; and second, by operationalising this support through an auditable, phase-gated agentic workflow. In this sense, the merging plan is not a replacement of one model by another, but a structured integration of three layers: ENISA as the normative-methodological reference, the AI-enhanced lifecycle as the conceptual extension, and the agentic workflow as the implementation mechanism.

This integration is necessary because the three elements solve different, but closely related, problems. ENISA provides a robust lifecycle structure and clear planning logic, including explicit go/no-go points, disciplined artefact production, and a strong emphasis on evaluation and improvement. The AI-enhanced lifecycle proposed in this dissertation builds on this structure by showing where AI can reduce workload and improve consistency without undermining human responsibility. The agentic workflow then translates this conceptual AI support into a controlled operational process in which specialised agents generate structured outputs, pass them through validation gates, and preserve full traceability of decisions and artefact evolution. The result is a merged model that remains aligned with European best practice while also addressing one of the key practical problems identified in this dissertation: the labour-intensive, uneven, and often weakly standardised preparation of cyber exercises across organizations and sectors.

The core principle of the merging plan is therefore simple: the ENISA phases remain unchanged as the authoritative lifecycle logic, while the AI-enhanced lifecycle and the agentic workflow are embedded into these phases as enabling layers. This means that the dissertation does not propose a competing lifecycle. Instead, it proposes an ENISA-aligned lifecycle that is made more operational, more scalable, and more auditable through structured AI augmentation. This is fully consistent with the dissertation's existing position that the resulting model should be understood as an ENISA-aligned cybersecurity exercise lifecycle with increased operational granularity and an AI augmentation concept that improves efficiency and learning impact without compromising governance requirements.

From a structural perspective, the merging plan can be understood as a three-level architecture.

More precisely:

- **Methodology layer = ENISA lifecycle backbone**
- **Augmentation layer = the AI support layer**
- **Implementation layer = the agentic workflow that realises that support layer**

In the three-level architecture, the supporting layer corresponds to the second level, namely the augmentation layer. This layer introduces the controlled AI support concept across the ENISA lifecycle phases. The third level, the implementation layer, does not constitute the support layer itself; rather, it operationalises that support layer through the agentic workflow and its specialised agents.

At the first level stands the **methodology layer**. This is the ENISA Cyber Security Exercise Methodology, which defines the six main phases of the exercise cycle: Initiation, Design, Preparation, Execution, Evaluation, and Moving Forward. These phases provide the stable

sequence, the expected logic of progression, and the minimum governance expectations for a high-quality cyber exercise. They also define the natural points for decision-making and review, especially through ENISA's emphasis on phase transitions and readiness checkpoints.

At the second level stands the augmentation layer. This is the dissertation's AI-enhanced lifecycle model, which introduces a controlled AI support layer across the ENISA phases. Here, AI is not treated as an autonomous planner or exercise controller, but as a bounded support capability. Its role is to consolidate evidence, draft artefacts, check consistency, structure scenario material, support evaluation design, assist evidence synthesis, and improve follow-through in remediation and re-testing. Across all phases, the same rule applies: AI recommends, humans decide. This preserves accountability while still allowing measurable efficiency gains.

The supporting layer is the dissertation's governed AI-assistance concept: a structured, non-autonomous layer that helps execute lifecycle tasks more consistently, measurably, and scalably, while humans retain methodological and operational control.

At the third level stands the **implementation layer**. This is the agentic workflow described in Chapter 5. The workflow decomposes the exercise-planning problem into bounded tasks performed by specialised agents, including the Requirements and Compliance Agent, Scope and Stakeholder Agent, Objectives and KPIs Agent, Scenario Architect Agent, MSEL/Inject Engineer Agent, and Evaluation Agent, all coordinated by an Orchestrator or Phase Controller. Each agent receives explicit inputs, produces structured outputs, and operates under phase-specific go/no-go gates. This design transforms the AI support layer from a general idea into a reproducible and testable mechanism.

The practical effect of this merging plan is that each ENISA phase becomes associated with two integrated questions: first, what must be achieved according to ENISA; and second, how can the dissertation's AI-enhanced and agentic model support this phase in a controlled and auditable way? This creates a direct mapping between recognised methodology and technical implementation.

In the **Initiation** phase, ENISA requires the exercise rationale, type, feasibility, and stakeholder

Agentic Workflow for Cyber Exercises

In this dissertation, the AI support layer refers to a bounded augmentation layer embedded across the cyber exercise lifecycle. Its purpose is to assist human planners and evaluators in repetitive, structure-dependent, and cross-check-intensive tasks by producing reviewable artefacts, consistency checks, and decision-support outputs. It does not replace the lifecycle methodology, nor does it transfer decision authority to AI. Instead, it defines where and under which governance conditions AI may support the lifecycle in a traceable, auditable, and human-controlled manner. The agentic workflow operationalises this support layer by assigning these bounded functions to specialised agents under phase-gated orchestration and human approval.

landscape to be defined. In the merged model, this remains unchanged at the methodological level. The added value comes from the AI and agentic layers: the Requirements and Compliance Agent extracts legal, regulatory, and operational constraints; the Scope and Stakeholder Agent defines in-scope and out-of-scope boundaries and the initial participation model; and the Orchestrator compiles these into the first version of the exercise plan and enforces the initiation gate. AI therefore

strengthens evidence-based framing, but the human exercise owner remains responsible for approving the purpose, scope, and feasibility baseline.

In the **Design** phase, ENISA requires the translation of intent into measurable objectives, refined scope, governance structure, and an early evaluation logic. In the merged model, this phase is operationalized by the Objectives and KPIs Agent, the refined Scope and Stakeholder Agent, and the Evaluation Agent. Their outputs ensure that objectives become measurable, that roles and decision levels are aligned with those objectives, and that an evaluation skeleton already exists before scenario development begins. This is one of the most important integration points, because it creates the traceable linkage that later connects objectives to injects, evidence sources, and findings.

In the **Preparation** phase, ENISA emphasizes scenario engineering, MSEL construction, inject design, artefact preparation, and evaluation-by-design. In the merged model, this is where the Scenario Architect Agent and the MSEL/Inject Engineer Agent become central. AI can generate scenario variants, propose storyline phases, produce inject drafts across channels, and check pacing and objective coverage. At the same time, evaluation artefacts such as observer sheets, questionnaires, and collection logic can be drafted in parallel. However, the dissertation's model explicitly requires human realism review, version control, and formal validation before these materials are accepted for execution. This prevents the workflow from drifting into unreviewed synthetic content production.

In the **Execution** phase, ENISA's priority is disciplined orchestration, active monitoring, and structured evidence capture. In the merged model, AI is deliberately kept in a supporting role. It can summarize exercise developments, track inject delivery, highlight pacing problems, structure event logs, and, where relevant, suggest adaptive branch injects. However, all real control actions remain with the human facilitation team. This is a critical safeguard in the merging plan. It ensures that AI is used to reduce cognitive load and improve situational awareness, not to displace facilitator judgement in live exercise management. The rule "AI recommends, humans decide" is therefore especially important in this phase.

In the **Evaluation phase**, ENISA requires defensible analysis and the production of an actionable After-Action Report. In the merged model, the Evaluation Agent assists in turning heterogeneous evidence into a coherent timeline, clustering observations, calculating time- and process-related indicators, and drafting findings in a structured format such as observation, impact, cause, and recommendation. Yet the methodological boundary remains explicit: claims, interpretations, and ratings must be reviewed and validated by human evaluators. This is essential both for scientific rigour and for practical credibility in regulated and safety-critical contexts.

Finally, in **Moving Forward**, ENISA frames improvement and re-testing as the endpoint of exercise value. In the merged model, this is operationalized by turning findings into a structured action backlog, linking tasks to ownership, deadlines, and evidence requirements, and proposing focused re-tests for later cycles. This is where the merged model connects most directly to the dissertation's broader argument about national cyber resilience: exercises should not end with lessons identified, but should create a trackable and auditable improvement loop. In other words, the merging plan ensures that the lifecycle, the AI support layer, and the workflow all converge toward measurable capability improvement rather than isolated exercise delivery.

Methodologically, the merging plan also clarifies the relationship between Chapter 4 and Chapter 5 of the dissertation. Chapter 4 should be read as defining the **reference lifecycle and its AI-enabled extension**, while Chapter 5 should be read as providing the **operational realisation** of that extension through a bounded agentic architecture. This division of roles is important for conceptual clarity. Without it, the reader may perceive the lifecycle model and the agentic workflow as separate contributions. In fact, they are part of one integrated design. The lifecycle explains where support is needed across the exercise phases; the workflow explains how this support can be delivered in a structured and auditable manner.

This integration also strengthens the dissertation's contribution after the publication of the ENISA methodology. Instead of weakening the originality of the lifecycle chapter, ENISA's release increases the practical relevance of the work. By aligning the proposed lifecycle to ENISA, the dissertation improves compatibility with a recognised European methodology and makes its results easier to adopt in practice. Originality then shifts from "creating another lifecycle" to "extending a recognised lifecycle with an auditable AI support layer and a concrete agentic implementation model." This is a stronger contribution, because it combines methodological legitimacy with operational innovation.

From a governance perspective, the merging plan also fits the dissertation's wider emphasis on repeatability, comparability, and accountability. ENISA provides the lifecycle logic; the AI-enhanced layer adds structured efficiency and evidence support; and the agentic workflow ensures that artefacts are produced in a stable schema, with explicit inputs, outputs, and approval points. This is particularly relevant for national and sectoral exercise programmes, where exercises must increasingly serve not only for learning, but also for defensible reporting, improvement tracking, and longitudinal comparison across cycles.

In summary, the concrete merging plan proposed in this dissertation can be expressed as follows: the ENISA Cyber Security Exercise Methodology remains the authoritative lifecycle structure; the dissertation's AI-enhanced lifecycle adds a controlled support layer across all phases; and the Chapter 5 agentic workflow provides the technical and procedural means to operationalize this support under explicit human control. This produces a single integrated model rather than three parallel ones. The resulting architecture is ENISA-aligned, AI-augmented, and workflow-driven. It is intended to preserve methodological rigour while making cyber exercise planning and evaluation more scalable, more consistent, and more auditable across organizations, sectors, and national contexts.

Practical integration rule set

To make the merged model actionable, the following integration rules apply throughout the lifecycle:

First, ENISA defines the mandatory phase structure and the expected core deliverables.

Second, AI support may only be used within clearly bounded tasks such as consolidation, drafting, consistency checking, and evidence structuring.

Third, each AI-supported step must be linked to provenance information, including inputs, versioning, timestamps, and approval status.

Fourth, every phase ends with a human-led go/no-go decision before the workflow proceeds.

Fifth, exercise control during execution remains human-led even where AI provides recommendations.

Sixth, evaluation outputs may be AI-assisted in structure, but never accepted without human validation.

Seventh, improvement measures must be linked back to evidence and forward to re-testing, so that the merged model supports a full capability cycle rather than isolated exercise delivery.

These rules ensure that the merging of ENISA, the AI-enhanced lifecycle, and the agentic workflow remains defensible in both scientific and operational terms.

The scientific contribution of Part III is an Auditable Agentic Workflow that operationalises lifecycle-level AI augmentation through specialised, bounded AI-supported responsibilities, structured artefacts, human approval gates, provenance, and policy constraints. Its novelty lies not in applying generative AI to cyber exercise planning, but in transforming AI assistance into a reproducible, testable, and governance-compatible organizational process.

Scientific Contribution of Part III

The principal scientific contribution of Part III is the development and practical implementation of the Auditable Agentic AI Workflow (C3). The research moves beyond conceptual architecture by operationalising three specialised AI agents within a sequential, RAG-supported exercise preparation process, incorporating defined responsibilities, structured inputs and outputs, mandatory human approval gates, and traceability mechanisms. The implementation demonstrates that selected cyber exercise preparation tasks can be organised into a functioning, human-supervised agentic workflow, while its comparative performance against conventional planning remains a distinct empirical evaluation question.

The third scientific contribution (C3) is the development and implementation of an Auditable Agentic AI Workflow for cyber exercise preparation, addressing Hypothesis H3 and Main Research Question MRQ3. The research extends beyond conceptual workflow design by implementing a functioning demonstrator comprising three specialised AI agents: the Objectives & KPIs Agent, Scenario Architect Agent, and MSEL / Inject Engineer Agent. These components operate within a sequential, retrieval-augmented workflow with defined responsibilities, structured inputs and outputs, source grounding, mandatory human approval gates, and traceability mechanisms.

The originality of C3 lies in transforming generative AI assistance into a structured, reproducible, human-supervised, and auditable cyber exercise engineering process. The implemented demonstrator operationalises selected activities within the Initiation, Design, and Preparation phases of the ENISA Cybersecurity Exercise Methodology. Its scientific contribution comprises both the implemented workflow architecture and the associated quantitative evaluation framework for examining whether agentic AI-supported preparation improves efficiency, exercise quality, internal consistency, and traceability compared with conventional human-only planning

Research Results and Interpretation

Validation examines whether objectives link correctly to scenarios, injects map to objectives and expected evidence, evaluation findings trace to observations, regulatory or policy claims are source-grounded, and required human approvals are recorded.
A five-point evaluation scale can be used, where 1 represents very poor performance and 5 represents excellent performance. For hallucination risk, the direction should be made explicit so that a higher score consistently represents better performance, for example by assessing 'hallucination control' or 'absence of unsupported claims'.
The Agentic Workflow is validated at artefact and workflow level rather than through a universal claim that AI improves national resilience. Principal criteria are accuracy, completeness, relevance, internal consistency, source grounding, traceability, provenance, human control, governance compatibility, hallucination control, and preparation efficiency.

Agentic Workflow Prototype

The main contribution of the prototype is not conversational interaction as such, but the design and evaluation of a controlled multi-agent workflow that supports the early phases of the ENISA-aligned cyber exercise lifecycle. In particular, the prototype focuses on assisting Initiation, Design, and Preparation by generating structured planning artefacts, including scope elements, objectives, KPIs, scenario components, and injects, under explicit human review and approval.
The contribution is not “a chatbot for cyber exercises.”
The central experimental focus of the prototype lies in testing whether a bounded multi-agent workflow can support the generation of exercise design artefacts and inject packages for the first three lifecycle phases in a traceable, auditable, and human-controlled manner.

Validation Framework

At exercise level, provenance is linked to the evidence chain developed in Part II: requirement → objective → scenario → inject → evidence → finding → corrective action → re-test. The organizational workflow therefore preserves both artefact provenance and exercise evidence provenance.
Every material artefact should preserve a trace from source to approved output. The minimum provenance chain is: source → AI-supported task → generated draft → reviewer action → approved artefact → version record → downstream use. This enables planners, evaluators, auditors, or governance bodies to reconstruct why an artefact exists and how it was transformed.

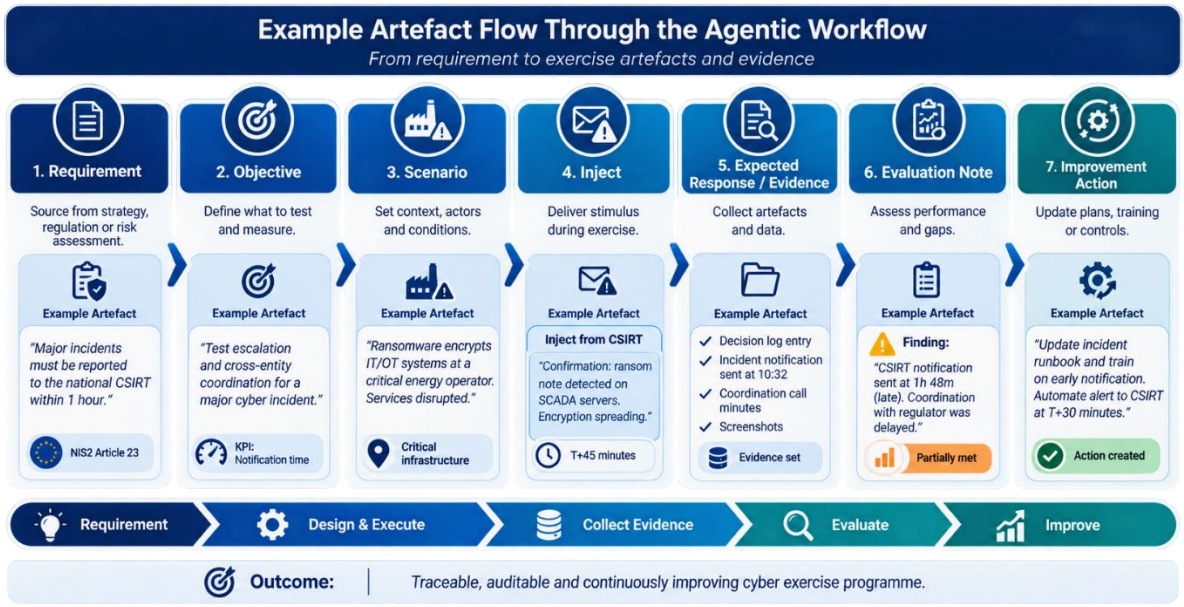


TABLE 14 - AGENTIC WORKLFOW EXAMPLE ARTEFACT

Provenance and Audit Architecture

The operational sequence is: AI-supported production → human review → accept / revise / reject → provenance log → progression. Where an artefact is rejected, it may be revised manually or regenerated through a bounded retry using explicit reviewer feedback. The AI system is never the approving authority.
Human control is applied proportionately. Low-risk drafting and structuring tasks may be AI-supported with routine review. Objectives, scenarios, injects, evidence mappings, and evaluation drafts require mandatory validation. Strategic exercise purpose, sensitive scenario approval, live exercise control, final findings, legal interpretation, remediation prioritisation, and residual-risk acceptance remain human decisions.

Human-in-the-Loop Control Architecture

3.2.1.7 *Implementation*

In course of my dissertation, I developed a local implementation of the above described concept. The contribution is a bounded agentic support system for the early cyber-exercise lifecycle, especially for the first three phases, where the workload is highest and much of the output is still manual, inconsistent, and hard to scale.

The main contribution of the proposed prototype lies not in the development of a chatbot interface, but in the **design of a bounded, local multi-agent workflow that supports the most preparation-intensive stages of the cyber exercise lifecycle**. The system focuses on the **first three phases - Initiation, Design, and Preparation** - and is intended to assist human planners in generating structured artefacts such as scope definitions, stakeholder mappings, objectives, KPIs,

scenario elements, and inject drafts. Particular emphasis is placed on inject generation, as this represents one of the most labour-intensive and quality-sensitive parts of exercise preparation. The prototype therefore serves as a research instrument for testing whether agentic support can improve consistency, traceability, and efficiency without reducing human control or accountability.

Scenario and Inject Engineering

Threat-actor modelling should also account for actors whose institutional affiliation is uncertain. Hacktivist groups, proxies, and cyber volunteers can operate in the space between independent activism and state-supported activity, creating ambiguity that is particularly suitable for testing escalation, attribution, and decision-making during exercises (Krasznay, 2024).[112]

This part of the dissertation builds on my practical expertise in organizing and delivering organizational cybersecurity exercises, as well as my involvement in the Austrian national KSÖ Planspiele. It gives an empirical, practitioner-informed description of the injects used in such exercises:

- Emails & Communication,
- Technical Data & Event,
- Videos & Cut scenes,
- Role-Player Interactions,
- Artifacts & Information,
- Business & Operational Impact and
- Live Technical Attacks.

Emails & Communication injects: Emails & Communication injects are messages that simulate the way information, uncertainty, and pressure enter an organization during a cyber incident. They are typically delivered through channels that participants use in daily operations, email, chat (e.g., Teams or Slack), ticketing systems, hotline calls, or SMS-style alerts, so that the exercise feels operationally realistic and not "separated" from normal communication patterns. The main purpose of this inject type is to test whether teams can recognise early warning signs, prioritise and triage incoming information, and coordinate across functions while the situation is still unclear. Typical examples include simulated phishing or spear-phishing emails (with safe links or attachments), internal threads reporting anomalies ("users are locked out," "unusual password reset requests," "payment confirmations failing"), executive messages asking for immediate status updates, customer or supplier inquiries, requests from legal or the data protection officer for facts and timelines, and media or regulator messages that create external urgency.

These injects are designed to drive concrete actions: participants should determine whether the message indicates a security incident or a general IT problem, route it through the correct escalation path, and ensure that the right roles are informed at the right time (e.g., SOC/IT operations, incident lead, CISO, crisis management, legal, communications, service owners). They also test the quality of operational communication: whether teams can produce short, accurate updates, avoid speculation, and maintain consistent messaging across multiple audiences despite incomplete evidence. Importantly, email and communication injects often reveal coordination weaknesses that purely technical injects do not, for example, duplicated work because teams are not aligned, delays because escalation thresholds are unclear, or conflicting statements to

[112] Krasznay, C. (2024). "Hacktivists, Proxy Groups, Cyber Volunteers." Academic and Applied Research in Military and Public Management Science, 23(3), 107–124.

management, customers, and staff. As exercise outputs, they provide strong evaluation evidence such as speed and correctness of triage and escalation, clarity and accuracy of written and verbal updates, appropriateness of stakeholder handling, and the completeness of documentation (what was known, when it was known, who approved which decision, and what was communicated externally and internally).

Technical Data & Event injects provide participants with the kinds of technical signals and operational facts that teams would normally rely on when investigating a suspected incident. They are used to create a realistic “picture on the ground” and to force participants to work with evidence, not assumptions. In practice, these injects can include SIEM alerts, EDR detections, authentication and identity logs, firewall/WAF events, DNS and proxy logs, cloud audit logs, vulnerability scan results, service desk tickets, or short forensic extracts. They are often presented with deliberate imperfections ,missing fields, time delays, duplicates, or conflicting indicators ,because that reflects how real investigations unfold. The injects are designed to prompt scoping (which systems and users are affected), correlation (which events are connected), and judgement (what is signal and what is noise). Typical participant actions include triage and classification, building and refining a working hypothesis about the attacker’s path, selecting containment measures that match the level of confidence, and documenting findings in a way that supports later review. From an evaluation perspective, technical-data injects provide clear observables: how quickly teams identify the relevant evidence, whether they correlate information across sources, whether their decisions are proportionate to uncertainty, and whether they maintain a consistent, shared understanding across technical and management roles.

Videos & Cut scenes injects are short, scripted multimedia elements that introduce context and pressure in a way that a log file cannot. They are especially useful in tabletop and hybrid exercises, where the scenario needs to “move” and participants must respond to external narratives. These injects can take the form of a simulated news report, a recorded executive message, a press conference excerpt, a social media clip, or a narrated timeline update that advances the scenario to a new stage. Their function is to shape situational awareness, add urgency, and introduce constraints such as reputational impact, public attention, political sensitivity, or regulatory scrutiny. They are most effective when they create a decision point: whether to communicate publicly, what to tell customers, whether to activate crisis management, or how to brief executives with limited technical certainty. Participants are expected to interpret the new context, align internal messaging, update leadership briefings, and translate strategic direction into practical priorities. The evaluation value is typically seen in the clarity and consistency of decisions, the quality of briefings and messaging, and the ability to manage pressure without losing control of the response process.

Role-Player Interactions injects are delivered through live interaction with facilitators or designated role-players who impersonate the stakeholders that responders have to deal with in real incidents. This can include executive leadership, legal counsel, data protection officers, communications/PR, HR, business service owners, key suppliers, customers, regulators, law enforcement, or journalists. The main purpose is to test cross-functional coordination and governance under realistic organizational dynamics, including conflicting objectives and incomplete information. Role-players can ask uncomfortable questions, demand timelines, challenge the chosen approach, or escalate external pressure (“Are you sure no data was leaked?”, “Do we need to notify today?”, “Can we keep the service online?”). Participants are expected to manage these interactions professionally, provide structured updates, ask for and record approvals, and ensure decisions remain traceable and aligned with policy and legal requirements. These injects create strong evaluation evidence around stakeholder management, escalation discipline, decision justification, communication quality, and the organization’s ability to keep a coherent command structure when priorities collide.

Artifacts & Information injects provide discrete pieces of information that participants must interpret, validate, and turn into action. Unlike continuous technical telemetry, these injects are often "objects" such as a threat intelligence report, a set of IOCs, a malware analysis summary, a screenshot of a ransom note, a synthetic sample of leaked data, an architecture diagram, an incident response playbook excerpt, or contractual clauses related to SLAs and breach notification obligations. Their purpose is to test information handling: source credibility, relevance, and correct interpretation. Well-designed artefacts include ambiguity or contradictions, because real incidents often involve partial or misleading information. Participants are expected to check plausibility, map the artefact to their working hypothesis, decide what additional evidence is needed, and update containment, recovery, and communication plans accordingly. From an evaluation point of view, artefacts are valuable because they make decision-making traceable: facilitators can see whether participants used the available documentation correctly, whether they identified red herrings, and whether they grounded their actions in verifiable inputs rather than instinct.

Threat-intelligence artefacts can provide a particularly valuable basis for exercise injects because they combine technical indicators with contextual information about threat actors, campaigns, tactics, and likely targets. Krasznay and Gyebnár (2021), in their analysis of cyber threat intelligence in energy systems, demonstrate both the potential and the limitations of CTI in supporting operational cybersecurity decisions. In exercise design, this means that intelligence injects should not be treated as unquestionable ground truth; their completeness, reliability, timeliness, and operational relevance should themselves become part of the decision-making challenge.[113]

Business & Operational Impact injects translate a technical situation into concrete consequences for the organization. They are used to ensure that the exercise tests resilience and governance, not just technical response. These injects can include customer complaints, revenue loss estimates, SLA breach warnings, disrupted business processes (payments, logistics, production), unavailability of critical applications, dependency failures due to third parties, or ,in OT settings ,potential safety implications and operational constraints. Their main function is to force prioritisation: which services must be restored first, which workarounds are acceptable, what risks can be temporarily accepted, and how to communicate service status to customers, partners, and authorities. They also test the integration of incident response with crisis management, business continuity, and disaster recovery processes. Participants are expected to assess impact and dependencies, define recovery priorities, allocate resources, approve controlled degradation or fallback processes, and keep leadership informed with clear, comparable impact reporting. The evaluation evidence includes whether priorities match critical services, whether decisions align with risk appetite and obligations, and whether technical and business teams maintain a shared recovery plan rather than working in parallel silos.

Live Technical Attacks injects take place in real time in a controlled environment such as a cyber range, lab, or isolated tenant, where an emulated adversary activity produces observable effects that defenders must detect and respond to. These injects can include safe execution of common attacker techniques ,credential abuse, lateral movement, privilege escalation, persistence, data staging, simulated exfiltration, or ransomware-like encryption ,always limited to non-production systems and designed to avoid harmful outcomes. Their purpose is to test operational capability under realistic tempo: monitoring, detection, triage, containment, eradication, and recovery, together with coordination and documentation. Participants are expected to use their tools (SIEM, EDR, SOAR, ticketing), isolate affected assets, block malicious activity, rotate credentials, collect forensic material, restore services, and verify that the threat is removed. These injects generate

[113] Krasznay, C., & Gyebnár, G. (2021). "Possibilities and Limitations of Cyber Threat Intelligence in Energy Systems." In 13th International Conference on Cyber Conflict: Going Viral (pp. 171–188). NATO CCDCOE.

measurable evidence such as time-to-detect and time-to-contain, but they also reveal qualitative strengths and weaknesses: how well teams follow runbooks, whether actions are properly documented for after-action review, whether coordination between roles holds under pressure, and whether recovery is verified rather than assumed.

Supporting Exercise Technologies and Environments

Cyber exercises are often introduced through their scenarios, roles, and facilitation techniques. In practice, the *technology layer* determines what an exercise can realistically test, how safely it can be executed, and what kind of evidence it can produce for evaluation and programme learning. The same scenario will lead to very different outcomes depending on whether it is run in a fully instrumented cyber range, in a simulated enterprise environment, via adversary-emulation tooling, or as a discussion-based tabletop. For national and defence-oriented programmes ,where exercises increasingly support resilience targets, assurance expectations, and longitudinal capability development ,technology choice is therefore not a secondary operational detail. It is a design decision that affects validity, repeatability, auditability, and the workload placed on facilitators.

“Cyber exercise technologies” should be understood as an ecosystem of enabling components rather than as one platform type. Mature programmes combine multiple technology categories because cyber crises are socio-technical: technical response actions must be coordinated with governance decisions, communications, legal obligations, and service continuity priorities. The categories below are structured from infrastructure-heavy, high-fidelity environments to facilitation-centric formats. They are not mutually exclusive, and they should be selected based on exercise objectives and participant maturity. A central programme design principle follows from this: high technical fidelity is not inherently superior if the objective is executive decision-making and cross-entity coordination; conversely, low-tech formats cannot credibly validate technical detection, containment, and recovery performance.

3.2.1.8 *Cyber Ranges as Integrated Exercise Environments*

A **cyber range** is an isolated, configurable environment that supports realistic technical exercises through controllable infrastructure and instrumentation. Typical range capabilities include virtualised networks and hosts, enterprise services (e.g., directory services, email, web applications), identity and access components, monitoring pipelines, and mechanisms for scenario control and reset. What distinguishes a cyber range from an ad-hoc lab is its ability to be *reproducible and observable*: environments can be instantiated from baselines, exercises can be re-run, and evaluation can draw on time-synchronised evidence.

In real-world practice, cyber ranges often underpin large-scale technical defence exercises. A well-known example is **Locked Shields**, which operates as a high-intensity blue-team defence exercise where participants must protect and maintain services within a simulated infrastructure under continuous pressure. Even when teams participate remotely, a central range environment ensures consistent technical ground truth, controlled attack stimuli, and comparable scoring and evidence collection. This illustrates a key governance advantage: in a national programme context, a range creates a stable technical reference that enables re-testing across cycles, which is essential for demonstrating improvement rather than repeating “lessons identified.”

Cyber ranges vary widely in fidelity. At lower fidelity, they may emulate a small enterprise network with minimal dependencies ,useful for learning triage, basic containment, and incident workflows. At higher fidelity, ranges can simulate enterprise complexity with segmented networks, multi-tier applications, backup/restore constraints, and monitoring stacks. In mission-critical environments, higher fidelity also means realistic constraints: limited administrative privileges, strict change-control assumptions, service continuity requirements, and, where relevant, OT/ICS safety

boundaries. These constraints prevent unrealistic “textbook fixes” and force teams to practice decision-making under the same trade-offs they would face in reality.

The limitations of cyber ranges are equally important. They require ongoing maintenance, skilled controllers, and significant preparation time. If the environment is overly complex for participants, exercises can degrade into troubleshooting infrastructure rather than validating capabilities. At programme level, comparability is also not automatic: different range providers can produce materially different experiences even if both claim “enterprise realism.” A national programme that uses ranges at scale therefore benefits from explicit environment baselines (minimum services, telemetry expectations, reset logic) and from documented assumptions that make results interpretable.

3.2.1.9 *Simulation, Emulation, and Digital Twins*

Beyond integrated ranges, many exercises rely on **simulation** or **emulation** approaches. These are conceptually distinct and support different learning goals. Simulation models behaviour abstractly, often focusing on consequences rather than technical artefacts ,e.g., “a ransomware event reduces service capacity by X% and triggers cascading impacts.” Emulation recreates representative systems and protocols in a way that allows realistic interaction ,e.g., an emulated OT network segment that supports meaningful response actions without exposing real operational systems. Digital twins aim to represent the configuration and dependency structure of real environments more closely, often used in mission-critical contexts where constraints define what “valid response” looks like.

Real-world crisis-management exercises often apply simulated impacts while maintaining functional realism. A national crisis exercise may not recreate the entire technical system, but it can realistically simulate degraded service states, public and political pressure, supply-chain effects, and regulatory reporting deadlines. In such settings, the technical cause is less important than the tested behaviours: escalation decisions, coordination across entities, prioritisation of services, and communication discipline. Digital twin approaches become particularly useful when decisions depend on real dependencies (e.g., segmented networks, shared identity services, constrained maintenance windows). The trade-off is cost and confidentiality: the closer the twin resembles reality, the more sensitive its representation becomes and the more effort is required to maintain it.

3.2.1.10 *Adversary Emulation and Attack-Orchestration Tooling*

A distinct technology category, sometimes embedded in ranges but conceptually separate, is **adversary emulation**: tooling that creates repeatable “attack stimuli” to trigger detection and response. This includes scripted attack chains, incident replays, and TTP-based emulation mapped to known attacker behaviour patterns. The purpose in an exercise context is not to teach exploitation; it is to generate controlled technical and informational events that require response action.

Many operational blue-team exercises rely on a controlled attack package that is executed by controllers. The advantage is repeatability: if the stimulus is consistent and logged, performance differences over time become interpretable. For example, a national programme may run a baseline emulation of a credential compromise leading to lateral movement and data exfiltration signals. In an early iteration, teams may detect late and respond inconsistently. In later cycles, the same baseline stimulus can be used to measure improvement in detection speed, containment decision quality, and recovery planning. To avoid predictability, variation can be introduced through parameterised changes ,alternate entry points, adjusted timing, partial telemetry degradation, or different secondary effects ,while preserving a stable core baseline.

In defence-oriented contexts, governance control is critical. Rules of engagement must prevent destructive actions, the executed chain must be versioned, and controller interventions must be logged to preserve auditability. Without such controls, adversary emulation can drift into unsafe practices or create exercise outcomes that cannot be defended because assumptions are undocumented.

3.2.1.11 *Technical Skills Platforms and Hands-On Lab Environments*

Between full cyber ranges and competitive formats sits a wide class of **hands-on lab and training platforms**. These environments provide guided technical tasks and structured learning tracks, often with automated validation and progress measurement. They are widely used for workforce development, SOC training, and pre-exercise preparation because they scale well and require fewer resources than full-scale simulations.

A typical real-world use pattern is layered preparation: teams first train on labs that focus on log analysis, triage workflows, detection rule tuning, and forensic fundamentals, then participate in a larger exercise where these skills are applied under pressure and in coordination with governance and communications. From a national programme perspective, this category supports a maturity progression strategy: lower-maturity organizations can build baseline competence before being expected to participate in high-intensity, multi-entity exercises.

The limitation is that lab environments generally simplify organizational reality. They often do not represent handover problems, the political and legal constraints of crisis governance, multi-stakeholder coordination, or real service-continuity trade-offs. For national preparedness validation, labs are therefore complementary ,useful for competence building but insufficient as primary evidence of organizational readiness.

3.2.1.12 *Competitive Formats: Capture-the-Flag and Attack–Defence Platforms*

Capture-the-Flag (CTF) and competition systems are ubiquitous in cybersecurity training ecosystems and appear in many national talent programmes and academic pipelines. They range from puzzle-based tasks to attack–defence competitions that require simultaneous offensive and defensive actions. Their main strength is motivation and scale: they provide measurable performance indicators and can engage large numbers of participants.

Real-world examples include national or regional CTF competitions used to identify talent for national cyber teams or to build a pipeline into more structured training. Many organizations also use internal CTFs for skills development and team engagement. However, competitive incentive structures can distort behaviours toward point optimisation rather than realistic incident response, documentation discipline, or safe operations. In defence-oriented readiness programmes, CTFs are therefore best used as feeder and training instruments rather than as the main validation layer. They are highly valuable to raise competence, but their outcomes should not be overinterpreted as evidence of organizational preparedness in real incident contexts.

3.2.1.13 *Exercise Management, MSEL, and Inject Orchestration Systems*

Beyond the technical simulation itself, large and complex exercises rely heavily on **exercise management technologies**. These systems support planning and controlling the exercise flow: timeline design, MSEL management, inject scheduling, role-player scripts, multi-channel delivery (email, chat, portals), and controller coordination dashboards. In distributed exercises, they become essential because controllers must manage multiple participants, maintain consistent pacing, and coordinate role-player actions without losing situational overview.

A real-world example is a multi-organization coordination exercise where participants receive injects via different channels: a simulated regulator request for incident status, a supplier message

about service degradation, a “leaked” document, and a technical indicator package from a SOC feed. Exercise management technology can coordinate delivery, track acknowledgements, and log controller decisions to adjust timing if participants fall behind or move faster than expected. From an auditability perspective, such tooling contributes a decisive capability: it preserves a trace of *what was delivered, when, and to whom* ,a prerequisite for credible evaluation and for resolving disputes about whether a response failure was due to participant action or exercise delivery issues.

The main limitation is rigidity. If orchestration systems force strict scripts, they can undermine adaptive facilitation, which is often necessary for realism. Mature use therefore combines automation with controlled flexibility, where controller interventions are logged and become part of the evaluation record.

3.2.1.14 *Observability, Telemetry, and Evidence Capture Technologies*

For exercises that are expected to be auditable and comparable, **evidence capture** is often more important than scenario creativity. Evidence capture includes log aggregation, endpoint and network telemetry, time synchronisation, structured observer note systems, screen recording in controlled settings (where appropriate), and repositories for storing artefacts. In governance-focused exercises, evidence capture can be centred on decision logs, situation reports, draft communications, and escalation records. In technical exercises, it expands to telemetry streams and system-state snapshots.

A common failure mode is either under-instrumentation ,leading to subjective after-action assessments ,or over-instrumentation ,leading to overwhelming data volumes that are not mapped to objectives. The design requirement is objective-linked observability. If an objective concerns escalation discipline, evidence capture should focus on timestamps, decision thresholds, and communication artefacts. If an objective concerns recovery capability, evidence capture should include restore actions, service health indicators, and recovery decision logs. Programme-level comparability depends on consistent evidence planning: without it, improvements cannot be tracked reliably across cycles.

3.2.1.15 *Crisis Communication and Coordination Platforms*

Cyber incidents trigger communication-intensive workflows: internal reporting, executive briefings, regulator notifications, stakeholder coordination, and sometimes public messaging. Many exercises therefore incorporate **crisis communication and coordination technologies**, such as secure messaging, incident ticketing, situation report templates, collaborative crisis rooms, and simulated media or stakeholder channels. The value of these tools lies in making socio-technical processes observable and reviewable.

In large coordination exercises, a typical inject is a request for a situation report from a national coordination body or a sector regulator, with strict time constraints and partial information. Participants must coordinate internally to assemble facts, decide what can be shared, align on messaging, and provide a coherent response. Technology supports this by providing traceable workflows: who drafted and approved the report, which information sources were used, and how updates were communicated. For defence-oriented programmes, such traceability is crucial because it enables evaluation of governance performance without exposing sensitive technical details.

3.2.1.16 *Tabletop Exercises and Facilitation-Centric Formats*

Tabletop exercises (TTX) remain a foundational format because they are accessible, scalable, and inclusive for non-technical stakeholders. Their technology can be minimal (slides and printed injects) or supported by structured collaboration tooling (shared canvases, decision trackers, polling, observation templates). Tabletop formats are particularly effective for validating escalation

paths, roles and responsibilities, decision authority, legal and regulatory obligations, and cross-entity coordination mechanisms.

A typical real-world example is an executive tabletop where a simulated ransomware incident creates service disruption and reputational pressure. Injects may include a media inquiry, a regulator reporting deadline, a supplier outage, and conflicting internal information about data exposure. The technical details remain abstract, but the exercise tests whether governance mechanisms function: who declares an incident, who communicates externally, how priorities are set for restoration, and how uncertainty is managed. Tabletop exercises are also frequently used as preparation for higher-fidelity exercises, ensuring that governance and communication routines are established before costly technical environments are employed.

The limitation is that technical performance cannot be validated. Therefore, for national programme comparability, tabletops require disciplined artefact structures and evaluation rubrics; otherwise, they produce unstructured narratives that are difficult to integrate into a capability cycle.

Sequencing and programme-level technology strategy

A national defence-oriented cyber exercise programme benefits from treating these technology categories as a **progression ladder** rather than as competing options. Early maturity and broad participation are supported by tabletop formats and coordination tooling. Functional exercise management and inject orchestration then enable process rehearsal under tempo, generating more consistent evidence. Cyber ranges, adversary emulation, and robust telemetry support technical validation and objective-linked measurement. At the highest maturity, emulation and digital twins increase realism for mission-critical contexts where constraints define validity. Across all levels, evidence capture and structured artefact management remain cross-cutting enablers, because without them exercises cannot support auditable evaluation or verified improvement across cycles.

By structuring cyber exercise technologies into distinct types and clarifying their contribution, limitations, and real-world application patterns, this chapter provides a basis for objective-driven technology selection in cyber exercise design. It supports the dissertation’s broader argument that repeatability and accountability do not emerge automatically from running more exercises; they are engineered through explicit governance, structured artefacts, and technology choices that make performance observable and comparable over time.

3.3 Prototype / Demonstrator Architecture

The purpose of the demonstrator is to operationalise a bounded subset of the proposed Auditable Agentic Workflow and to provide a practical basis for evaluating Hypothesis H3. The demonstrator is not intended to implement the complete cyber exercise lifecycle, nor to replace experienced exercise planners. Instead, it focuses on a deliberately limited sequence of planning activities in which structured AI support can be observed, measured, and compared with conventional manual preparation.

This design follows the central principle of the dissertation that AI support should be introduced through specialised, bounded tasks with clearly defined inputs, outputs, and human approval points. The full conceptual workflow contains seven components: the Orchestrator / Phase Controller, Requirements & Compliance Agent, Scope & Stakeholder Agent, Objectives & KPIs Agent, Scenario Architect Agent, MSEL / Inject Engineer Agent, and Evaluation Agent. The

dissertation defines these components as part of a sequential and auditable workflow in which artefacts produced by one component become inputs for subsequent activities.

For the demonstrator, however, only three specialised components are implemented:

1. **Objectives & KPIs Agent**
2. **Scenario Architect Agent**
3. **MSEL / Inject Engineer Agent**

This reduced implementation was selected because it captures one of the most important logical chains in cyber exercise engineering:

Exercise Objectives → Scenario → Injects → Observable Evidence

The demonstrator therefore represents a functional subset of the complete architecture rather than a competing or simplified conceptual model. Its purpose is to test whether selected parts of the proposed agentic workflow can be technically operationalised and evaluated under controlled conditions.

The demonstrator belongs to the organizational and operational research layer of the dissertation.

The application of artificial intelligence in military and national-security environments must be examined in relation to operational requirements, institutional responsibilities, and the broader consequences of technological adoption. The multidisciplinary volume edited by Kenedli and Magyar (2025) provides a relevant Hungarian military-scientific context for these **considerations by addressing AI and its implications across several professional domain**s.[114] The present research contributes to this field through the development, implementation, and empirical evaluation of an Auditable Agentic AI Workflow (C3). Rather than treating AI as an unrestricted general-purpose assistant, the proposed approach integrates specialised agents, retrieval-augmented organisational knowledge, structured exercise artefacts, and mandatory human approval into a controlled cyber exercise preparation process.

The international and European layer establishes the external methodological and regulatory baseline. The national layer translates this baseline into a governance model and introduces the ENISA-based AI-enhanced cyber exercise lifecycle. The organizational layer subsequently operationalises selected AI augmentation points through specialised AI-supported components. The dissertation explicitly distinguishes between the lifecycle, which defines where AI support may be used and under which governance conditions, and the agentic workflow, which defines how such support can be technically implemented.

The demonstrator therefore represents the transition from conceptual model to technical operationalisation:

ENISA Cybersecurity Exercise Methodology

- ✓ **AI-Enhanced Cyber Exercise Lifecycle**
- ✓ **Auditable Agentic Workflow**
- ✓ **Bounded Demonstrator**
- ✓ **Structured Validation**

[114] Kenedli, T., & Magyar, S. (Eds.). (2025). A mesterséges intelligencia hatásainak specifikus vizsgálata. Katonai Nemzetbiztonsági Szolgálat. https://knbsz.gov.hu/kiadvanyok

Its primary scientific function is not to demonstrate autonomous exercise generation, but to provide an observable implementation through which the characteristics proposed in H3 can be examined.

The demonstrator is implemented in **two technical environments**. The **first implementation runs locally on a dedicated Mac Mini** and serves as the controlled research environment for development, configuration, repeated testing, logging, and comparison of the manual and AI-assisted workflows. In this setup, LM Studio provides the local language model runtime, AnythingLLM provides retrieval-augmented access to the approved organisational knowledge base, and OpenClaw is used to implement the specialised agent roles and the sequential workflow. This local environment allows the model configuration, prompts, retrieval settings, input documents, outputs, and human approval decisions to remain under direct control and supports reproducible experimentation without requiring external cloud processing.

The second implementation is deployed in the AI Lab of the University of St. Pölten (USTP) and is used as a controlled testing and educational environment with students. The purpose of this environment is not to introduce a different scientific model, but to instantiate the same bounded demonstrator in a multi-user laboratory setting. Students can interact with the same underlying workflow and work on the same type of structured cyber exercise artefacts under predefined instructions and supervision. This makes it possible to examine the usability and practical comprehensibility of the workflow beyond the researcher's own local environment, while still maintaining the same architectural logic, fixed use case, predefined agent responsibilities, and human-control principles. The local Mac Mini environment therefore functions primarily as the research and validation platform, while the USTP AI Lab functions as the controlled user-testing and teaching environment.

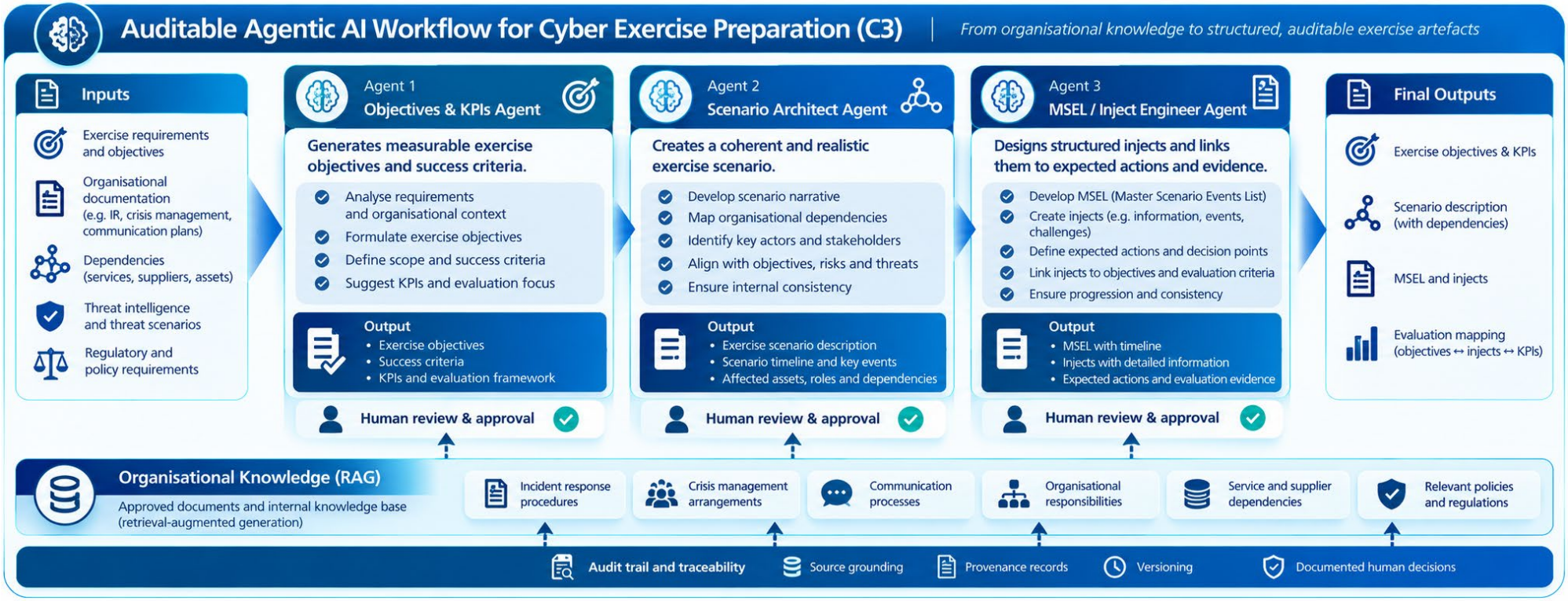


FIGURE 26 - AGENTIC AI WORKFLOW

The figure illustrates the proposed Auditable Agentic AI Workflow (C3), demonstrating how organisational knowledge is transformed into structured, traceable, and human-approved cyber exercise artefacts. The architecture consists of three sequentially connected AI agents: the Objectives & KPIs Agent, which defines measurable exercise objectives, success criteria, and evaluation indicators; the Scenario Architect Agent, which develops coherent exercise scenarios incorporating organisational dependencies, stakeholders, and relevant threats; and the MSEL / Inject Engineer Agent, which generates structured exercise injects linked to expected participant actions and evaluation evidence.

All three agents are supported by a shared Retrieval-Augmented Generation (RAG) knowledge base containing approved organisational documentation, including incident response plans, crisis management procedures, communication arrangements, responsibilities, and service dependencies. Mandatory human review and approval gates separate the agent stages, ensuring that generated outputs are validated before further processing. A cross-cutting auditability layer maintains source grounding, provenance records, versioning, and documented human decisions. The workflow produces measurable objectives and KPIs, scenario documentation, a Master Scenario Events List (MSEL), and explicit evaluation mappings. The architecture demonstrates the operationalisation of bounded, accountable, and auditable AI assistance within the Initiation, Design, and Preparation phases of the cyber exercise lifecycle, while preserving human decision-making authority throughout the process.

In both environments, the organisational documents are stored in the retrieval environment and are queried as needed by the agents. Each agent receives a predefined responsibility, input structure, output format, and behavioural constraints. Human approval is mandatory between the stages, meaning that an artefact can only progress after it has been reviewed and accepted by the human exercise planner or, in the educational setting, by the designated reviewer or instructor. Rejected outputs can either be corrected manually or returned for bounded regeneration. The AI therefore supports drafting, structuring, and consistency checking but does not hold approval authority. Each experimental run records the model and prompt version, input artefacts, retrieved sources, generated outputs, timestamps, and the human approval decision, enabling traceability and reproducibility in line with the dissertation's validation approach.

Empirical Validation of the Agentic AI Cyber Exercise Demonstrator

To ensure comparability between runs and between the two implementation environments, the demonstrator **uses the same synthetic defence-sector use case** (scenario) and the same organisational dataset in every execution. The modelled organisation is a fictional Defence Logistics and Communications Support Command (DLCSC) responsible for non-classified logistics coordination, transport and supply scheduling, maintenance support, administrative IT services, internal communications, and coordination with external IT and logistics providers. No classified information, real military unit, operational plan, real defence infrastructure, weapons system, intelligence platform, or operational command-and-control environment is included.

The scenario models a medium-complexity cyber incident originating from the compromise of an external IT maintenance provider. Credentials associated with the supplier are used to gain access to a remote-management environment connected to selected non-classified support systems. The attacker subsequently establishes persistence, abuses privileged service accounts, disrupts the logistics scheduling environment, deploys ransomware to selected application and file servers, and exfiltrates a limited amount of non-classified logistics and personnel-support information. During the incident, internal communication services also become partially degraded, while uncertainty remains as to whether additional systems may be affected. Although classified military systems are explicitly excluded, the disruption begins to affect logistics coordination, maintenance scheduling, supply-status visibility, and communication between support functions. The scenario therefore creates operational pressure without modelling tactical military cyber operations. Its progression can be summarised as third-party compromise → privileged-account misuse → service disruption → ransomware deployment → limited data exfiltration → communication degradation.

The exercise request remains identical in every run: the DLCSC must be provided with a two-hour cyber exercise designed to evaluate its ability to detect and escalate a significant cyber incident, coordinate technical and management response, maintain critical logistics-support functions under degraded conditions, manage a compromised third-party provider, preserve relevant evidence, and establish recovery priorities during a ransomware-related incident. The baseline knowledge set includes a synthetic organisation profile, incident response plan, crisis-management and escalation procedure, critical-service and dependency register, third-party provider description, exercise requirements, and evaluation requirements. Keeping these inputs fixed ensures that differences in the outputs can be attributed primarily to the preparation method and, where relevant, to the user interaction context rather than to changes in scenario or organisational content.

The first component, the Objectives & KPIs Agent, transforms the approved organisational context and exercise requirements into a small set of measurable objectives. Its output links each objective to a success criterion or KPI, an expected evidence source, and a source or rationale. The agent

is not allowed to generate scenario content or injects. After human approval, the Scenario Architect Agent receives the approved objectives and develops the scenario backbone, including the initial situation, incident progression, affected functions, dependencies, and decision points. The scenario must remain aligned with the approved objectives and may not redefine the exercise purpose. Once the scenario is approved, the MSEL / Inject Engineer Agent transforms it into approximately eight to ten structured injects. Each inject contains an identifier, scenario time, target role, inject content, linked objective, expected participant action, expected evidence, and delivery method. This creates an explicit traceability chain from Objective → Scenario Event → Inject → Expected Action → Evidence.

Human control is retained throughout the entire process. The operational sequence is AI-supported production → Human review → Accept / Revise / Reject → Provenance record → Progression. No AI-generated objective, scenario, or inject automatically becomes an approved exercise artefact. This directly implements the dissertation's governing principle that AI recommends, humans decide, and is consistent with the requirement that consequential artefacts remain subject to mandatory human validation.

The demonstrator supports a controlled comparison between a manual and an AI-assisted preparation condition. In the manual condition, a human planner produces the objectives, scenario backbone, and MSEL from the fixed input documents. In the AI-assisted condition, the same artefacts are produced through the three-agent workflow using the same source material and exercise request. The local Mac Mini implementation is used for repeatable experimental runs under tightly controlled technical conditions, while the USTP AI Lab implementation allows the same workflow to be tested with students in a supervised multi-user setting. The two environments therefore complement each other: the first supports technical reproducibility and controlled comparison, while the second provides additional evidence regarding practical usability, comprehensibility, and interaction with the workflow.

Evaluation focuses on preparation time, completeness of required artefact fields, objective coverage, inject-to-objective traceability, internal consistency across objectives, scenario and injects, the number of human corrections required, and structured expert assessment of artefact quality. In the USTP AI Lab environment, additional observations may be collected concerning usability, task clarity, perceived usefulness, and the extent to which users can understand and correctly apply the human-approval logic. These user-oriented observations should be treated as complementary evidence rather than as a replacement for the core artefact-based validation. The purpose of the evaluation is not to demonstrate that AI universally produces superior cyber exercises. Instead, the experiment tests whether bounded and source-grounded AI support can improve selected aspects of preparation efficiency, structured production, consistency, traceability, and practical usability while preserving human authority and accountability.

The demonstrator therefore provides the operational validation mechanism for H3. Bounded AI support is represented by three specialised components with narrowly defined responsibilities; human control is represented by mandatory approval gates; efficiency can be observed through preparation time; internal consistency can be examined through objective–scenario–inject relationships; traceability is represented by explicit mappings and recorded sources; and auditability is supported through prompt versions, artefact versions, timestamps, source records, and human approval decisions. H3 is thus translated from a conceptual proposition into a set of observable implementation properties.

The scientific value of the demonstrator does not lie in showing that a generative AI model can write exercise content. Its contribution is the operationalisation of AI assistance as a bounded, source-grounded, sequential, human-controlled, and auditable cyber exercise engineering

process. The use of two implementation environments strengthens this contribution by separating controlled technical validation from supervised user-oriented testing while maintaining the same underlying architecture and use case. In this way, the demonstrator creates a practical bridge between the ENISA-based AI-enhanced lifecycle and the organisational implementation layer of the dissertation, while providing a reproducible and practically testable environment for evaluating the benefits and limitations of responsible AI support in defence-oriented cyber exercise preparation.

Empirical Evaluation and Verification of Hypothesis H3

The validation of Hypothesis H3 was designed to examine whether a human-supervised, retrieval-augmented agentic AI workflow can improve the efficiency, quality, internal consistency, and traceability of cyber exercise preparation compared with conventional human-only planning. The evaluation focused on the Initiation, Design, and Preparation phases of the ENISA Cybersecurity Exercise Methodology, corresponding to the implemented scope of the demonstrator. The objective was to assess measurable differences in exercise-planning outputs rather than to evaluate autonomous exercise execution or operational military capabilities.

The demonstrator was implemented in two complementary environments: a local research environment using a Mac Mini with LM Studio, AnythingLLM, and OpenClaw, and a supervised educational environment at the University of St. Pölten AI Lab. Both environments used the same synthetic defence-oriented organisational context, based on the fictional Defence Logistics and Communications Support Command (DLCSC). The underlying knowledge base contained modelled incident-response procedures, escalation arrangements, organisational responsibilities, communication processes, service dependencies, and exercise requirements. Three specialised components: the Objectives & KPIs Agent, Scenario Architect Agent, and MSEL / Inject Engineer Agent, supported sequential artefact generation, with mandatory human approval between workflow stages.

The comparative evaluation design incorporated student exercises conducted across two academic semesters, with two classes per semester and four working groups per class. The design provided sixteen groups organised into eight scenario-matched comparisons between conventional planning and the agentic AI-supported approach. Participating groups were required to develop equivalent exercise deliverables under comparable scenario requirements. Evaluation criteria included total human planning effort, independently assessed exercise quality, internal consistency faults, and traceability between objectives, scenario events, injects, expected participant actions, and evaluation evidence. The comparison was designed to use matched observations and standardised assessment criteria to reduce the influence of differences in exercise complexity. A separate international practitioner survey involving 40 respondents provides a complementary source for examining professional expectations and the perceived relevance of AI-assisted exercise development, but is not included in the quantitative performance comparison presented below.

For methodological illustration, the dataset representing **eight matched group** comparisons was analysed. The results indicate that agentic AI-supported preparation required an average of 14.88 person-hours, compared with 19.91 person-hours for conventional planning, corresponding to an estimated reduction of 25.3%. The mean paired difference was 5.04 person-hours (95% CI: 4.24–5.84), with lower preparation effort in all eight comparisons. Exercise quality, evaluated using a standardised 100-point assessment framework, increased from an average of 72.75 to 79.88 points, representing a mean improvement of 7.13 points (95% CI: 3.43–10.82). Positive quality differences were observed in seven of eight matched comparisons.

The analysis also suggested improvements in artefact consistency and evidence traceability. The average number of internal consistency faults decreased from 9.25 to approximately 5.13 per 60

assessed items, representing a reduction of 44.6% and a mean paired difference of 4.13 faults (95% CI: 2.06–6.19). Traceability coverage increased from 69.6% to 88.9%, corresponding to an improvement of 19.25 percentage points (95% CI: 10.9–27.6). Both indicators showed favourable differences for the AI-supported condition in seven of eight matched comparisons. The confidence intervals were estimated from paired differences using a t-distribution-based approach.

H3 indicator	Average AI Advantage	95% confidence intervall
Efficiency	5,04 hours saved (25,3%)	4,24–5,84 hours
Exercise quality	+7,13 points / 100	+3,43 to +10,82 points
Internal consistency	4,13 fewer faults (44,6%)	2,06 to 6,19 faults
Traceability	+19,25 % points	+10,9 to +27,6 % point

TABLE 15 - IMPROVEMENT COMPARISON

Quantitative comparison of conventional and agentic AI-supported cyber exercise preparation across eight matched group comparisons. The table above presents mean paired differences, 95% confidence intervals, and the number of comparisons favouring the agentic AI-supported approach for planning efficiency, exercise quality, internal consistency, and traceability.

Taken together, the results demonstrate how H3 can be assessed through four complementary performance indicators. The observed pattern within the simulated dataset would be consistent with the proposition that bounded agentic AI support can reduce preparation effort while improving the structural quality and traceability of exercise artefacts.

The scope of the evaluation is deliberately restricted to exercise preparation. Although the conceptual agentic architecture addresses the complete ENISA exercise lifecycle, the implemented demonstrator focuses on objective formulation, scenario construction, and inject engineering. Consequently, the evaluation does not establish improvements in live exercise conduct, participant operational performance, incident recovery, or national cyber resilience. Its scientific contribution lies in establishing a reproducible, artefact-oriented validation approach through which the efficiency and governance compatibility of human-supervised agentic AI support can be assessed. The final assessment of H3 will depend on whether the actual experimental observations confirm the direction and magnitude of the improvements illustrated here.

3.4 Regulatory Pathway III – Organizational Compliance Operationalisation

At the organizational and operational level, the regulatory pathway translates external legal, regulatory, and national governance requirements into concrete controls within the Auditable Agentic Workflow and into observable exercise evidence. The purpose is not to automate legal interpretation or determine compliance through AI, but to ensure that approved organizational obligations are reflected consistently in exercise objectives, workflow constraints, scenario design, evidence requirements, and human approval mechanisms. In this sense, the workflow functions as an operationalisation layer between formal requirements and practical capability validation.

The pathway begins with the identification of applicable requirements from authoritative and approved sources. European and national obligations arising from instruments such as NIS2 and NISG 2026, CER and RKEG, DORA, GDPR, and the EU AI Act are interpreted within the organization and translated into operational expectations. This operationalisation challenge is particularly relevant in the context of NIS2, where practical implementation requires organizations to translate broad regulatory obligations into functioning governance, risk-management, incident-response, continuity, and reporting arrangements (Krasznay, 2024).[115]

These expectations may concern incident handling, escalation, reporting, continuity, third-party dependencies, management responsibility, data protection, human oversight, or resilience testing. Once approved, they can be represented within the workflow as bounded requirements, control conditions, exercise objectives, evidence fields, and decision gates. This creates a traceable relationship between the original requirement and the exercise artefacts generated later in the process.

The European digital regulatory environment increasingly consists of interconnected legal instruments that address different dimensions of the digital transformation, including cybersecurity, data protection, platform governance, artificial intelligence and market regulation. Tóth (2024)[116] describes this development as part of the broader regulatory response of the European Union to digital transformation, in which sector-specific and horizontal regulatory instruments increasingly interact. This broader regulatory context is relevant for the present dissertation because cyber exercise governance and AI-supported exercise processes do not operate within a single legal framework, but at the intersection of several European regulatory domains.

This approach builds on prior research into LLM-supported compliance operationalisation. In the CO2 (Co-Compliance Officer) methodology[117], demonstrated how regulatory and compliance knowledge can be represented through structured knowledge models and used to support AI-assisted compliance checking. The present dissertation transfers this principle to the cyber-exercise domain: rather than using AI to autonomously determine compliance, approved regulatory requirements are transformed into bounded workflow constraints, exercise objectives, evidence requirements, and human-controlled decision gates.

Within cybersecurity and incident governance, for example, requirements concerning reporting, escalation, continuity, supply-chain security, or management responsibility can be translated into scenario conditions and measurable exercise objectives. An exercise may test whether an incident is recognised and escalated correctly, whether required stakeholders are informed, whether continuity procedures are activated, or whether third-party dependencies are managed effectively. The resulting evidence may include reporting latency, completeness of decision logs, escalation correctness, coordination performance, and the quality of situational awareness. Such evidence does not prove legal compliance, but it provides structured information on whether the organizational mechanisms associated with those obligations function under simulated stress.

The same principle applies to critical-entity and continuity obligations. Requirements derived from CER and RKEG can be operationalized through scenarios involving service degradation, dependency failure, alternative procedures, crisis coordination, and recovery priorities. In financial-sector contexts, DORA-related resilience expectations can be translated into objectives concerning ICT disruption, third-party dependencies, recovery, and structured evidence collection. The

[115] Krasznay, C. (2024). “A NIS2 irányelv kihívásai és gyakorlati alkalmazása.” Védelem Tudomány, 2024 special issue, 1–5.

[116] Kozák, K. (2026). Tóth András – A digitális átalakulás szabályozása az Európai Unióban. In Medias Res, 15(1), 195–204.

[117] Turaga, V. S. P., Páhi, T., Tjoa, S., Siami-Namini, S., & Siami Namin, A. (2025). CO2 (Co-Compliance Officer): An LLM-based ontology-driven methodology for generating knowledge graphs and AI compliance checking. IEEE Access, 13, 210576–210606.

workflow can support these activities while remaining clearly distinct from formal supervisory testing or TLPT processes.

Data protection and AI governance operate as cross-cutting constraints across the workflow. GDPR and national data-protection requirements influence how participant data, exercise logs, communications, and AI inputs are processed. Accordingly, the workflow should apply data minimisation, purpose limitation, access control, retention rules, and approved processing environments. AI governance requirements further reinforce the need for bounded agent responsibilities, source grounding, explicit human approval gates, policy constraints, audit logging, and retained human decision authority. AI-supported components may assist with analysis, drafting, consistency checking, and evidence structuring, but they do not independently determine legal interpretations, compliance status, or binding organizational decisions.

The organizational regulatory pathway therefore culminates in the use of cyber exercises as a form of **compliance-relevant operational validation**. Exercises can demonstrate whether selected processes and controls function effectively under realistic conditions, but they do not substitute for legal assessment, certification, supervision, or formal compliance determination. Their value lies in making organizational behaviour observable and in generating structured evidence that can be linked to findings, corrective actions, and subsequent re-testing.

The pathway can therefore be understood as a continuous chain from external requirement to operational evidence: approved regulatory obligations are translated into workflow controls and exercise objectives, these are represented through scenarios and injects, execution generates evidence, evaluation produces findings, and findings are converted into corrective actions that may later be re-tested. In this way, Regulatory Pathway III completes the cross-layer logic of the dissertation by connecting European and national requirements with organizational implementation, controlled AI support, and evidence-based improvement.

3.5 Assessment of H3 and Part III Summary

Part III addressed how the AI augmentation defined at lifecycle level can be implemented within organizations through a controlled workflow. The proposed Agentic Workflow decomposes exercise planning and evaluation into bounded responsibilities, structured artefacts, explicit human approval points, source-grounded tasks, provenance records, and policy constraints.
Within the scope of artefact- and workflow-level validation, H3 is supported as a design proposition. A controlled Agentic Workflow provides a more reproducible and auditable basis for AI-supported exercise production than open-ended use of a general-purpose assistant because it makes dependencies, review decisions, and provenance explicit.
The dissertation does not claim that the workflow autonomously improves national resilience or replaces expert planners. Its validated contribution is the organizational architecture through which AI support can improve consistency, traceability, structured production, and reviewability while preserving human decision authority.

Taken together with Part II, the result is a clear separation of responsibilities: ENISA defines the exercise process; the AI-enhanced lifecycle defines where AI may support that process; and the Auditable Agentic Workflow defines how such support can be operationalized under organizational governance controls.

4 Conclusion and Scientific Contributions

4.1 Integrated Answer to the Research Problem

This dissertation examined how cyber exercises can evolve from heterogeneous training activities into repeatable, comparable, auditable, and improvement-oriented instruments of national cyber

resilience. The research addressed this problem through three connected layers: international and European convergence, national governance, and organizational AI-supported operationalization. The integrated finding is that national cyber exercise effectiveness does not depend primarily on conducting more exercises. It depends on whether exercises are connected through common governance, lifecycle, data, evidence, and improvement structures. International methodologies and European regulation increasingly create a shared direction; national governance is required to translate that direction into a coherent capability system; and organizational implementation determines whether intended controls and evidence structures function in practice.

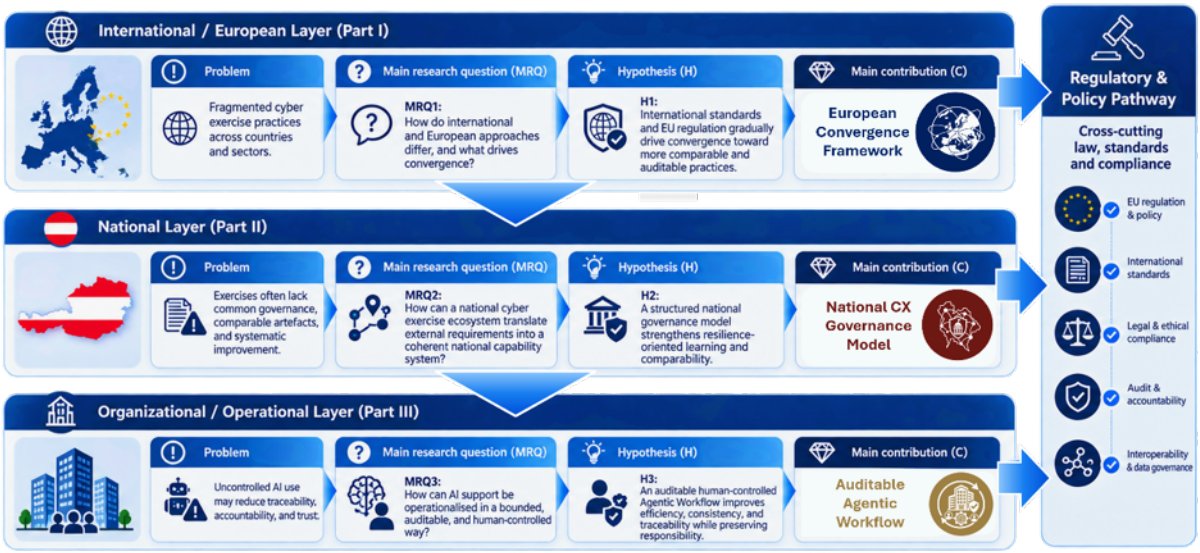


FIGURE 27 - THREE LAYERS AND CONTRIBUTIONS

4.2 H1 – International and European Convergence

H1 proposed that the international and European cyber exercise landscape remains fragmented while standards, methodologies, regulatory requirements, and policy developments increasingly create convergence toward more documented, comparable, auditable, and resilience-oriented practice.

The analysis supports H1 with an important qualification: convergence does not mean uniformity. Scenarios, technologies, sectors, maturity levels, and institutional arrangements remain heterogeneous. Convergence is more visible in underlying quality and governance principles: objective-driven planning, lifecycle orientation, defined responsibilities, structured artefacts, evidence-based evaluation, dependency awareness, corrective actions, and continuous improvement.

The scientific contribution associated with H1 is the European Convergence Framework. It converts a broad observation of convergence into a structured analytical model and an emerging minimum baseline that can be used to compare heterogeneous exercise practices without requiring identical exercise formats.

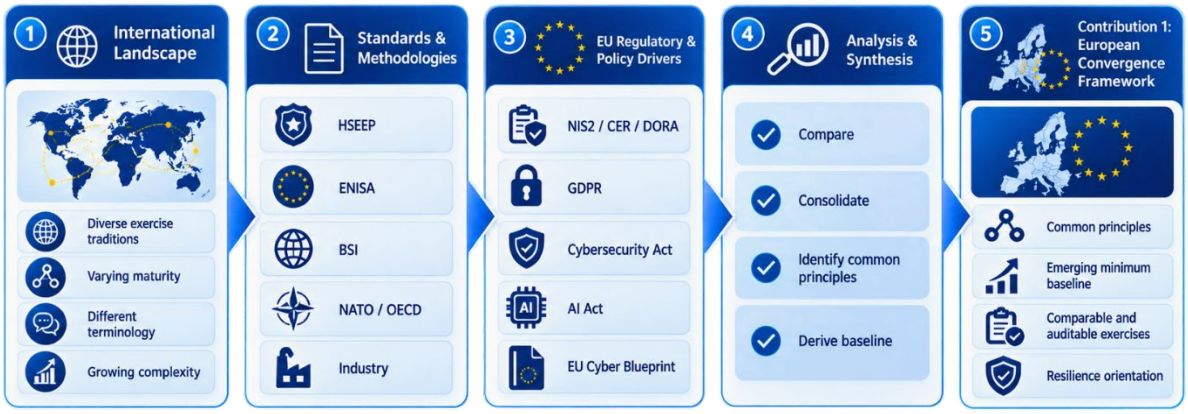


FIGURE 28 – (H1) FROM FRAGMENTED PRACTICE TO A EUROPEAN BASELINE

4.3 H2 – National Cyber Exercise Governance

H2 proposed that a national cyber exercise ecosystem based on defined governance responsibilities, common quality criteria, comparable artefacts and datasets, coordinated programmes, a structured lifecycle, and mandatory improvement loops provides a more suitable basis for resilience-oriented improvement than fragmented or ad hoc exercise activity.

The Austrian reference case demonstrates both the availability of important national capabilities and the governance challenge created by distributed responsibilities, multiple sectors, confidentiality constraints, and changing institutional structures. The National Cyber Exercise Governance Model addresses this challenge by adding a programme layer connecting strategic objectives, exercise portfolio management, taxonomy, minimum datasets, common artefacts, KPIs, evaluation, corrective actions, and re-testing.

The ENISA-based AI-Enhanced Cyber Exercise Lifecycle complements the governance model by providing a common methodological process. The lifecycle retains ENISA's six phases and extends them with bounded AI augmentation, human approval, provenance, evidence traceability, and controlled progression. H2 is therefore supported as a constructive governance proposition within the scope of model coherence, applicability, comparability, and improvement logic.

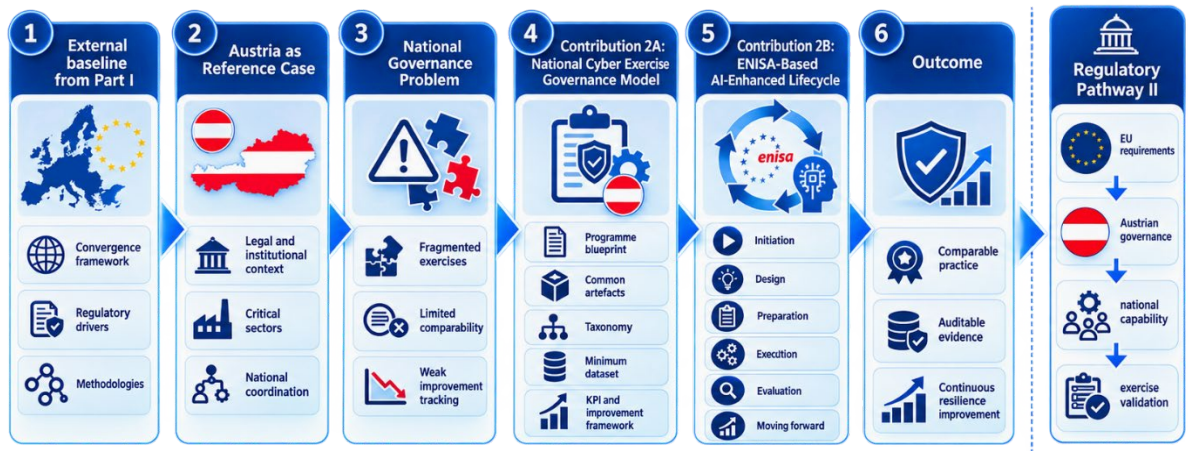


FIGURE 29 - FROM EUROPEAN CONVERGENCE TO NATIONAL CX GOVERNANCE

4.4 H3 – Auditable AI Operationalization

Hypothesis H3 states that operationalising AI support through a bounded, auditable, and human-controlled agentic workflow can improve the efficiency, internal consistency, traceability, and structured production of cyber exercise artefacts while preserving human responsibility, regulatory compliance, security, and accountability. The corresponding Main Research Question (MRQ3) examines how an auditable and human-controlled agentic workflow can operationalise AI support

across the cyber exercise lifecycle in a way that improves planning efficiency, consistency, and traceability while preserving accountability and manageable human oversight.

Hypothesis H3 and the corresponding research question MRQ3 are addressed through the development and implementation of a functioning, human-supervised agentic AI demonstrator for cyber exercise preparation. The implemented system comprises three specialised agents operating on a shared, retrieval-augmented organisational knowledge base and producing structured exercise objectives, scenario documentation, and MSEL injects under mandatory human review and approval. The demonstrator provides a concrete technical implementation of the proposed Auditable Agentic AI Workflow (C3), covering the Initiation, Design, and Preparation phases of the ENISA Cybersecurity Exercise Methodology. Its implementation supports the feasibility of the proposed architecture, while the extent of improvements in planning efficiency, artefact quality, internal consistency, and traceability must be determined from the actual comparative measurements.

The evaluation data, executed by students and other participants, indicated a 25.3% reduction in planning effort, a 7.13-point improvement in exercise quality, 44.6% fewer internal consistency faults, and an increase in traceability coverage from 69.6% to 88.9%. These results show the potential performance advantages of structured agentic AI support.

The main scientific contribution (C3) is the development of a reproducible, auditable, and human-controlled AI workflow for cyber exercise preparation, supported by a quantitative evaluation framework. The demonstrator provides the operational foundation for assessing H3 and addressing MRQ3 within the three implemented lifecycle phases, while the broader six-phase architecture remains a conceptual extension requiring further validation.

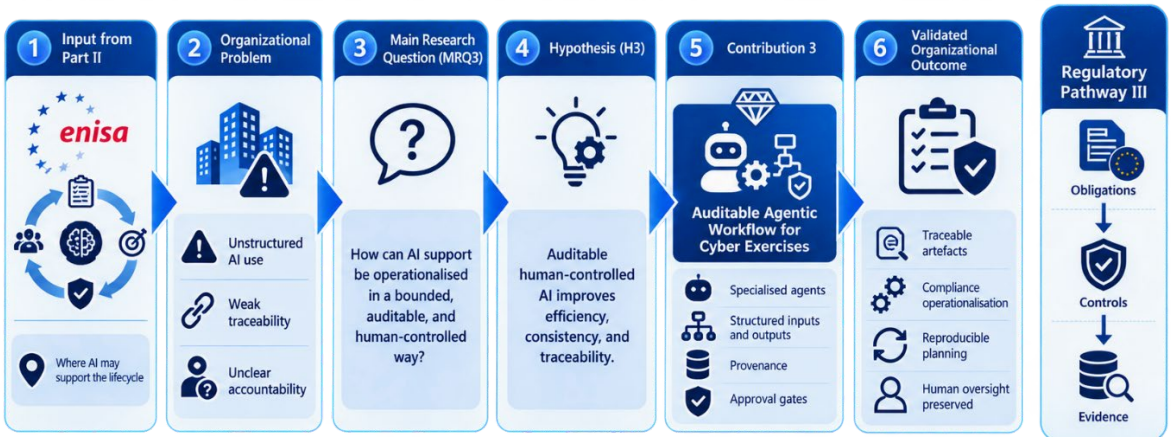


FIGURE 30 - FROM LIFECYCLE-LEVEL AI SUPPORT TO ORGANIZATIONAL IMPLEMENTATION

The scientific contribution of H3 (C3) is the development of an auditable, human-supervised agentic AI workflow that integrates retrieval-augmented generation, specialised AI agents, structured artefact production, and mandatory human approval gates into cyber exercise preparation. The research operationalises this approach through a defence-oriented demonstrator covering the Initiation, Design, and Preparation phases of the ENISA Cybersecurity Exercise Methodology and establishes a quantitative framework for evaluating planning efficiency, exercise quality, internal consistency, and traceability against conventional human-only approaches. Its originality lies in transforming generative AI assistance into a structured, reproducible, and governance-compatible process, while empirical validation of its performance advantages remains dependent on actual experimental measurements.

The military-scientific relevance of the dissertation can also be situated within contemporary Hungarian research on the implications of artificial intelligence for defence and national security.[118] **The dissertation extends this broader research context through three interconnected contributions: a European Convergence Framework for cyber exercise comparability (C1), a National Cyber Exercise Governance Model supporting coordinated and evidence-oriented capability development (C2), and an implemented and empirically evaluated Auditable Agentic AI Workflow for defence-oriented cyber exercise preparation (C3).**

From the military-scientific perspective, the dissertation complements existing Hungarian research on generative AI-supported national-security analysis and decision support (Nagy, 2025).[119] Its specific contribution is the implementation and empirical evaluation of a controlled agentic AI workflow for cyber exercise preparation, connecting AI-assisted analytical processes with measurable exercise objectives, structured scenario development, and auditable human decision-making.

4.5 Validation and Strength of Evidence

The dissertation does not claim universal empirical proof that the proposed models automatically increase national cyber resilience, reduce costs, or improve realism in every context. Validation focuses on properties that can be examined within the research scope: coherence, completeness, internal consistency, comparability, traceability, provenance, governance compatibility, preservation of human authority, structured evidence generation, and support for continuous improvement.
This delimitation is methodologically important. National resilience is influenced by many external factors and cannot be attributed to one exercise programme or AI workflow alone. The dissertation therefore validates the architecture and mechanisms through which resilience-oriented exercise learning can be made more systematic and auditable.

4.6 Limitations

The first limitation is the use of Austria as the principal national reference case. Austria provides a realistic and information-rich governance environment, but institutional structures, legal systems, military-civilian relationships, and exercise maturity differ across Member States. The proposed model is therefore designed to be adaptable rather than universally prescriptive.

A second limitation concerns access to sensitive exercise information. Detailed exercise artefacts, vulnerability findings, intelligence inputs, and regulatory testing results are often confidential. This limits empirical comparison and reinforces the need for a minimum-dataset approach based on controlled aggregation.

A third limitation concerns the rapid evolution of the legal and institutional environment. The Austrian NISG 2026 regime enters into operation on 1 October 2026, after the dissertation's August 2026 cut-off. The analysis therefore distinguishes between enacted law and operational implementation. Future research should evaluate how the new authority and coordination structures function in practice.

A further limitation concerns the scope and empirical evaluation of the implemented agentic AI demonstrator. Although the three-agent workflow has been developed and implemented in the local research environment and the University of St. Pölten AI Lab, its operational coverage is restricted to the Initiation, Design, and Preparation phases of the ENISA Cybersecurity Exercise

[118] Kenedli, T., & Magyar, S. (Eds.). (2025). A mesterséges intelligencia hatásainak specifikus vizsgálata. Katonai Nemzetbiztonsági Szolgálat. https://knbsz.gov.hu/kiadvanyok

[119] Nagy, V. (2025). A generatív mesterséges intelligencia felhasználásának területei a nemzetbiztonsági elemző-értékelő munkában, különös tekintettel a döntéstámogatásra. In T. Kenedli & S. Magyar (Eds.), *A mesterséges intelligencia hatásainak specifikus vizsgálata* (pp. 348–369). Katonai Nemzetbiztonsági Szolgálat. https://knbsz.gov.hu/kiadvanyok

Methodology. The implementation establishes an operational basis for assessing the proposed architecture but does not, by itself, establish superiority over conventional human-only exercise preparation.

4.7 Future Research

The research has developed and implemented a functioning three-agent demonstrator of the Auditable Agentic AI Workflow in a local research environment and a supervised university AI laboratory setting. Future research therefore does not primarily concern the initial implementation of the demonstrator, but its broader empirical validation, technical refinement, and extension to additional cyber exercise lifecycle activities.

A principal research priority is the analysis of actual comparative experimental observations to determine whether agentic AI-supported preparation produces statistically and practically meaningful improvements in planning efficiency, exercise quality, internal consistency, and traceability compared with conventional human-only approaches. Future evaluations should include larger and more diverse participant groups, independent expert assessment, standardised evaluation rubrics, and replication across different exercise scenarios and organisational contexts.

Further technical development should extend the current three-agent implementation beyond the Initiation, Design, and Preparation phases towards the Execution, Evaluation, and Moving Forward phases of the ENISA Cybersecurity Exercise Methodology. Additional research should examine the reliability of agent coordination, retrieval accuracy, provenance records, human approval mechanisms, error detection, and secure operation under more complex organisational requirements.

Finally, the demonstrator should be evaluated in representative defence-oriented and critical-infrastructure environments, subject to appropriate information-security, confidentiality, and governance controls. These investigations would establish the extent to which the implemented workflow can support more complex exercise programmes while maintaining human authority, traceability, accountability, and reproducibility.

4.8 Contributions and Final Conclusion

The dissertation has addressed a central challenge of contemporary cyber resilience: cyber exercises are increasingly expected to do more than provide training. They are expected to support preparedness, generate evidence, inform governance, contribute to continuous improvement, and operate within complex regulatory and AI-enabled environments. Yet the research has shown that the current cyber exercise landscape remains fragmented across countries, sectors, organisations, methodologies, artefacts, evaluation practices, and improvement mechanisms. The problem is therefore not the absence of cyber exercises, but the absence of a sufficiently integrated system through which heterogeneous exercise activities can become comparable, auditable, improvement-oriented, and responsibly scalable.

The dissertation responds to this problem through a cumulative three-layer research architecture, resulting in **three scientific contributions: the European Convergence Framework (C1), the National Cyber Exercise Governance Model (C2), and the Auditable Agentic AI Workflow (C3).** These contributions address different levels of the cyber exercise ecosystem while establishing a coherent relationship between international methodological principles, national governance, and organisational implementation.

At the international and European level, the research demonstrated that substantial diversity remains in exercise terminology, governance, exercise formats, lifecycle models, documentation, and evaluation practices. At the same time, recurring principles can be identified across established methodologies, selected European regulatory requirements, and policy developments. These recurring principles form the basis of the **European Convergence Framework (C1)**, which

identifies eight analytical dimensions: conceptual clarity, governance and accountability, lifecycle orientation, structured artefacts, evaluation and measurement, auditability and evidence, dependency and ecosystem management, and continuous improvement. A comparative evidence matrix was developed to distinguish explicit, partial, and insufficient documentary support across selected international methodologies and European regulatory instruments. The scientific contribution of C1 is therefore not a claim of complete European harmonisation or increasing convergence over time, but the systematic identification and consolidation of common methodological and governance principles that can support comparability across heterogeneous cyber exercise approaches.

At the national level, the research examined how this external baseline can be translated into a coherent cyber exercise governance system. Austria was selected as the principal reference case because its distributed institutional environment provides a relevant setting in which European requirements, national legislation, governmental institutions, CSIRTs, regulators, critical entities, public-private structures, and recurring exercise activities interact. The case analysis was informed by documentary evidence and practitioner-informed contextual knowledge, with the latter distinguished from independently validated empirical observations. The analysis highlighted the importance of coordination across institutional and sectoral boundaries, particularly where organisational responsibilities, dependencies, and exercise-related improvement processes intersect.

Building on this analysis, the dissertation developed a **Cyber Exercise Taxonomy and Minimum Dataset** and integrated them into the broader **National Cyber Exercise Governance Model (C2)** and Programme Blueprint. The taxonomy provides a common descriptive language, while the Minimum Dataset defines the core information required for portfolio visibility, comparability, evidence management, and longitudinal learning. The governance model connects strategic priorities, institutional responsibilities, exercise portfolios, standardised artefacts, performance measurement, corrective-action ownership, and subsequent verification. Its central scientific contribution is a structured national governance architecture designed to transform cyber exercises from isolated activities into coordinated elements of a sustained capability-development system.

A particularly important finding of Part II is that exercise value does not arise from exercise completion alone. An After-Action Report, lessons-identified document, or set of recommendations does not demonstrate that resilience has improved. Improvement becomes defensible only when exercise evidence is converted into findings, findings into corrective actions, corrective actions into implemented changes, and those changes into subsequent verification. The resulting logic can be expressed as a continuous sequence of **exercise, evidence, finding, corrective action, re-test, and verified improvement**. This improvement loop represents a central mechanism through which national cyber exercise governance can support evidence-based resilience development without reducing resilience to a single metric.

The ENISA-based AI-Enhanced Cyber Exercise Lifecycle establishes the methodological connection between the national governance model and the operational implementation of AI support. The dissertation deliberately adopts the six recognised ENISA phases of Initiation, Design, Preparation, Execution, Evaluation, and Moving Forward rather than introducing a competing lifecycle model. AI augmentation opportunities are mapped to these phases, while human authority, traceability, provenance, bounded AI use, data minimisation, and phase-gated progression are defined as cross-cutting requirements. The lifecycle therefore provides an organising structure for responsible AI integration and connects the governance principles developed in C2 with the operational mechanisms implemented in C3. It is treated as an enabling methodological component rather than an independent scientific contribution.



FIGURE 31 - SUMMARY OF CONTRIBUTIONS

The figure summarizes the three scientific contributions: European Convergence Framework (C1), National Cyber Exercise Governance Model (C2), and Auditable Agentic AI Workflow (C3). The ENISA-based AI-Enhanced Cyber Exercise Lifecycle provides the methodological connection between national governance and organisational AI operationalisation.

At the organisational and operational level, the dissertation develops the Auditable Agentic AI Workflow (C3) to operationalise AI support within a controlled and accountable cyber exercise preparation process. The workflow decomposes exercise development into specialised, bounded, and reviewable AI-supported tasks with defined inputs, structured outputs, retrieval-based source grounding, human approval gates, provenance records, and controlled progression. Its objective is not to replace human exercise planners, but to support time-intensive planning activities while preserving human decision-making authority, accountability, and governance compatibility.

The contribution is demonstrated through a defence-oriented implementation using a synthetic Defence Logistics and Communications Ransomware Exercise scenario. The implemented demonstrator contains three specialised components: the Objectives & KPIs Agent, Scenario Architect Agent, and MSEL / Inject Engineer Agent. These components sequentially generate measurable exercise objectives, a scenario aligned with organisational dependencies, and structured injects linked to expected actions and evaluation evidence. The demonstrator integrates retrieval-augmented generation with organisational reference documents and mandatory human approval points. Its operational scope is deliberately limited to the Initiation, Design, and Preparation phases of the ENISA methodology, whereas the remaining lifecycle phases are addressed conceptually.

To investigate the performance-related aspects of H3, a quantitative evaluation framework was developed to compare agentic AI-supported preparation with conventional human-only exercise planning. The evaluation design incorporates eight scenario-matched group comparisons across four university classes and focuses on four indicators: planning efficiency, exercise quality, internal consistency, and traceability. The test data demonstrated potential improvements of 25.3% in planning efficiency, 7.13 points in exercise quality, 44.6% fewer consistency faults, and an increase in traceability coverage from 69.6% to 88.9%. These values illustrate the proposed evaluation methodology and the types of measurable benefits that could support H3; they do not constitute verified empirical findings. Confirmation of these performance advantages requires analysis of the actual experimental observations, including their statistical uncertainty and methodological limitations.

The scientific added value of C3 is therefore not the demonstration that generative AI can create exercise content, since this capability is already available through general-purpose AI systems. Rather, it lies in transforming AI assistance into a reproducible, source-grounded, reviewable, and auditable engineering process suitable for governance-sensitive cyber exercise environments. By combining specialised agent responsibilities, structured artefact relationships, documented human approval, and measurable evaluation criteria, C3 provides both an operational implementation of responsible AI support and a methodological basis for assessing its performance against conventional approaches.

The cross-cutting regulatory and policy pathway connects all three research layers. European requirements and international methodologies establish external expectations; national legislation and governance structures translate these expectations into institutional and sectoral responsibilities; and organisations operationalise them through controls, exercise objectives, evidence requirements, and improvement mechanisms. Within this architecture, cyber exercises are not treated as automatic proof of legal compliance. Rather, they provide one mechanism through which operational capabilities associated with regulatory and preparedness requirements can be evaluated under realistic conditions. This distinction allows cyber exercises to contribute to assurance and resilience development without reducing them to compliance exercises.

Taken together, the three scientific contributions form one coherent research chain. The European Convergence Framework (C1) identifies common methodological and governance principles across heterogeneous exercise approaches. The National Cyber Exercise Governance Model (C2) translates these principles into a structured national capability-development architecture. The Auditable Agentic AI Workflow (C3) operationalises selected exercise-planning activities through bounded, human-supervised, and traceable AI support. The ENISA-based AI-enhanced lifecycle connects the national governance and operational research layers, while the regulatory pathway

links European requirements to national and organisational responsibilities. The improvement loop connects exercise evidence with corrective actions and subsequent verification.

The findings provide differentiated support for the three research hypotheses within the methodological boundaries of the dissertation. H1 is supported at the level of documentary analysis through the identification of recurring governance, lifecycle, documentation, evaluation, and improvement principles despite continuing fragmentation. H2 is supported as a constructive governance proposition through the development of a national model integrating common information structures, institutional responsibilities, programme coordination, measurement, and verified improvement; its real-world effectiveness remains to be established through subsequent implementation and longitudinal assessment. H3 is operationalised through a bounded, auditable, and human-controlled demonstrator and an associated comparative evaluation framework. While its architectural feasibility can be examined through the implemented workflow, the hypothesis of superior efficiency, quality, consistency, and traceability requires confirmation using actual experimental measurements.

These conclusions are intentionally bounded. The dissertation does not claim that the proposed models automatically increase national cyber resilience in every context, nor does it establish universal causal effects on preparedness, cost, or operational performance. National resilience depends on numerous external factors, including institutional maturity, resources, threat exposure, organisational culture, and technical capability. The contribution of this research is therefore primarily architectural, methodological, and operational: it defines and connects the structures, relationships, controls, artefacts, and evidence mechanisms through which cyber exercising can become more systematic, comparable, auditable, and improvement-oriented.

The broader scientific implication is that cyber exercises should be understood not merely as training events, but as instruments of governance, evidence generation, and capability assurance. Their strategic value increases when they are embedded within a system that connects policy, institutional responsibility, exercise design, performance evaluation, remediation, and re-testing. Under such conditions, cyber exercises can serve as a practical interface between formal preparedness requirements and observable organisational capabilities.

The final proposition of the dissertation can therefore be stated as follows: national cyber resilience becomes more governable and more systematically improvable when international and European methodological principles, national governance, organisational implementation, structured exercise evidence, continuous improvement, and responsible AI support are connected within one coherent cyber exercise ecosystem.

In this sense, the research moves the discussion beyond the question of whether organisations conduct cyber exercises. The more relevant question is whether those exercises form part of a system capable of learning from evidence, assigning responsibility, verifying change, and preserving accountability as complexity and automation increase. The dissertation demonstrates how such a system can be conceptually and operationally structured through the integration of convergence principles, national governance, common data and artefact logic, lifecycle discipline, human-controlled AI augmentation, and evidence-based improvement.

The central conclusion is therefore simple but consequential:

CYBER EXERCISES CONTRIBUTE MOST EFFECTIVELY TO NATIONAL RESILIENCE WHEN THEY ARE NOT TREATED AS ISOLATED EVENTS, BUT AS GOVERNED, EVIDENCE-DRIVEN, REPEATABLE, AND CONTINUOUSLY IMPROVING COMPONENTS OF A NATIONAL CAPABILITY SYSTEM.

5 Declaration of Generative AI Use and Author Responsibility

Generative Artificial Intelligence (AI) tools, including Large Language Models (LLMs) such as OpenAI's ChatGPT, were used in a supporting capacity during the preparation of this dissertation. Their use was primarily limited to linguistic refinement, reformulation and restructuring of selected text passages, improving academic clarity and consistency, and assisting in the creation and visual refinement of conceptual diagrams, illustrations, and research visualisations.

AI-generated outputs were treated as supporting materials rather than independent scientific sources or empirical evidence. Where synthetic data were used for illustrative statistical visualisations, these are explicitly identified as simulated examples and distinguished from empirical research findings.

The research objectives, methodological decisions, scientific contributions, interpretations, and conclusions remain the responsibility of the author. Generative AI was used as an auxiliary writing and visualisation tool and does not replace the author's independent scientific reasoning, critical assessment, or academic accountability.

6 Abbreviations

AAR	After Action Report
AI	Artificial Intelligence
AI-CX-LM	AI-Enhanced Cyber Exercise Lifecycle Model
BCM	Business Continuity Management
BCMS	Business Continuity Management System
BIA	Business Impact Analysis
BSI	Bundesamt für Sicherheit in der Informationstechnik (German Federal Office for Information Security)
CCDCOE	Cooperative Cyber Defence Centre of Excellence
CER	Critical Entities Resilience Directive
CERT	Computer Emergency Response Team
CISO	Chief Information Security Officer
CSIRT	Computer Security Incident Response Team
CTB	Cyber Battle Training
CTF	Capture the Flag
CX	Cyber Exercise
DORA	Digital Operational Resilience Act
EDR	Endpoint Detection and Response
ENISA	European Union Agency for Cybersecurity
EU	European Union
EU-CyCLONe	European Cyber Crisis Liaison Organization Network
GDPR	General Data Protection Regulation
H	Hypothesis
HR	Human Resources
HSEEP	Homeland Security Exercise and Evaluation Program
ICS	Industrial Control Systems
ICT	Information and Communication Technology
IOC	Indicator of Compromise
IP	Improvement Plan
IPCR	Integrated Political Crisis Response
ISMS	Information Security Management System
IT	Information Technology
KII	Critical Information Infrastructure
KPI	Key Performance Indicator
KSÖ	Kompetenzzentrum Sicheres Österreich (früher Kuratorium)
MRQ	Main Research Queastion

MSEL	Master Scenario Event List
NATO	North Atlantic Treaty Organization
NIS	Network and Information Security Directive
NIS2	Network and Information Systems Directive (recast)
OT/ICS	Operational Technology / Industrial Control Systems
PDCA	Plan–Do–Check–Act
PMO	Programme Management Office
PR	Public Relations
Q	Questions
RACI	Responsible, Accountable, Consulted, Informed
SIEM	Security Information and Event Management
SLA	Service Level Agreement
SOAR	Security Orchestration, Automation and Response
SOC	Security Operations Center
TIBER-EU	Threat Intelligence-Based Ethical Red Teaming
TLP	Traffic Light Protocol
TLPT	Threat-Led Penetration Test
TTX	Tabletop Exercise
AbwA	Abwehramt
AdD	Anbieter digitaler Dienste
AEC	Austrian Energy CERT
AHC	Austrian Health CERT
BK/C4	Bundeskriminalamt/Cyber Crime Competence Center
BKA	Bundeskanzleramt
BMEIA	Bundesministerium für europäische und internationale Angelegenheiten
BMI	Bundesministerium für Inneres
BMLV	Bundesministerium für Landesverteidigung
BwD	Betreiber wesentlicher Dienste
CERT.at	Nationales Computer-Notfallteam
CKM	Cyberkrisen-Management
CKM-KA	CKM-Koordinationsausschuss
CSIRTs Network	Europäisches Computer-Notfallteam Netzwerk
CSS	Cybersicherheit-Steuerungsgruppe

CSP	Cybersicherheit-Plattform
DSN	Direktion Staatsschutz und Nachrichtendienst
ECCC	Europäische Kompetenzgemeinschaft für Cybersicherheit
EdÖV	Einrichtungen der öffentlichen Verwaltung
EU-CyCLONe	European Cyber Crisis Liaison Organization Network
GovCERT	Computer-Notfallteam der öffentlichen Verwaltung
HNaA	Heeres-Nachrichtenamt
IKDOK	Innerer Kreis der Operativen Koordinierungsstruktur
MilCyZ	Militärisches Cyberzentrum
NIS	BMI, Abteilung IV/S/2
NIS-CG	NIS-Kooperationsgruppe
OpKoord	Operative Koordinierungsstruktur
SKKM	Staatliches Krisen- und Katastrophenschutzmanagement

7 Appendix A - Scientific Posters

National University of Public Service | Doctoral School of Military Sciences | 07. Theory of Defence Informatics and Communication

International Convergence, National Governance, and AI Support in Cyber Exercise Ecosystems

Timea Pahi , BSc, MA | Supervisor: Dr. András Tóth, PhD, Associate Professor

Research Problem

Cyber exercise practices remain fragmented across countries and organizations, limiting comparability, auditability, and systematic learning.
A coherent national framework is needed to connect governance, common standards, evidence, and continuous improvement.
At the same time, AI support must be integrated in a controlled, transparent, and human-supervised way.

Research Aim

to show how cyber exercises can evolve from isolated training events into repeatable, comparable, auditable, and AI-supported instruments for strengthening national cyber resilience.

Plans are worthless, but planning is everything.

Dwight D. Eisenhower

Cyber exercises turn planning into tested preparedness, and preparedness into national resilience.

Methodology & Three Analytical Layers

EUROPEAN LAYER	**Problem:** Cyber exercise practices are fragmented across countries and sectors in terminology, methodologies, governance, documentation, and evaluation. **H1:** Despite this fragmentation, international standards and European regulation are creating convergence towards more comparable, auditable, and resilience-oriented practices.	**MRQ1:** How do international and European approaches to cyber exercises differ, and which standards, regulatory requirements, and policy developments are driving convergence towards a more comparable and auditable cyber exercise baseline?	**C1:** Consolidated international and European baseline of methodologies, standards, regulatory and policy drivers; identification of fragmentation, convergence trends, and emerging minimum expectations.
NATIONAL LAYER	**Problem:** Individual and sectoral cyber exercises are often conducted separately, without a common national governance model, comparable data, or systematic improvement mechanisms. **H2:** A structured national cyber exercise ecosystem with defined roles, comparable lifecycle structures provides a stronger basis for national resilience than isolated or ad hoc exercises.	**MRQ2:** How can a national defence-oriented cyber exercise ecosystem be designed and governed so that it translates international and European requirements into comparable, auditable, and continuously improving national cyber exercise practice?	**C2:** National cyber exercise governance model; programme blueprint; taxonomy; minimum dataset; KPI and improvement framework; ENISA-based AI-enhanced cyber exercise lifecycle
ORGANIZATONAL LAYER	**Problem:** AI can support cyber exercise planning and evaluation, but uncontrolled use may reduce traceability, accountability, quality, and trust. **H3:**. A bounded, auditable, and human-controlled Agentic Workflow can improve efficiency, consistency, traceability, and artefact quality while preserving human responsibility.	**MRQ3:** How can an auditable and human-controlled agentic workflow operationalise AI support across the cyber exercise lifecycle in a way that improves planning efficiency, consistency, and traceability while preserving accountability and human oversight?	**C3:** Auditable Agentic Workflow, specialised AI-supported responsibilities, structured inputs and outputs, provenance; human approval gate, phase-gated execution, organizational implementation of lifecycle-level AI support.

REGULATORY & POLICY



FIGURE 32 - SCIENTIFIC POSTER – METHODOLOGY



FIGURE 33 - SCIENTIFIC POSTER OF FUTURE CX (HU)

8 Appendix B - Cyber Exercises and Formats

This appendix provides a structured list of widely referenced cyber exercise series and exercise formats that can serve as benchmarks when designing, governing, or evaluating a national cyber exercise programme. The purpose is not to rank exercises or to reproduce internal design details, but to indicate what each exercise family is typically used for and which transferable best practices can be derived for scalable, repeatable, and auditable exercising.

International and multinational flagship exercises. At the high end of technical intensity, large-scale blue-team defence exercises are commonly used to validate operational defence under sustained pressure, including service continuity, incident handling performance, and team tempo, often supported by scoring and strong instrumentation. Two frequently referenced examples are *NATO CCDCOE Locked Shields* (blue-team collective defence with service availability, incident response and crisis elements) and *NATO CCDCOE Crossed Swords* (red–blue dynamics designed to create adversarial pressure and operational learning). Complementing these technical formats, defence-oriented coordination exercises emphasise command-and-control interfaces, cross-entity coordination, escalation routines, and decision-making under uncertainty. A central multinational reference point in this category is *NATO Cyber Coalition*, which typically stresses multi-stakeholder coordination across a defence ecosystem. At European level, cross-border crisis coordination exercises focus on information sharing, shared situational awareness, and alignment of communications across many organizations. Two prominent examples are *ENISA Cyber Europe* (EU-wide crisis coordination and cooperation in a recurring large-scale format) and *ENISA Blue OLEx* (operational-level cooperation and response coordination).

National crisis-management and whole-of-society exercise formats. A second family of reference formats treats cyber incidents explicitly as national crises with cross-sector impacts. These exercises usually place their main emphasis on crisis governance, inter-ministerial coordination, strategic communication, and the continuity of critical state functions rather than on deep technical defence validation alone. A widely known European example is *LÜKEX (Germany)*, which operates as a national crisis-management exercise series with cross-sector and government coordination focus. Beyond branded series, many countries run national crisis-management exercises that integrate cyber as one dimension in broader “whole-of-government” or “whole-of-society” crisis scenarios; although these are not always internationally labelled, they represent structurally relevant reference formats for national governance design.

National cyber exercise programmes and recurring national series. A third category consists of recurring national cyber exercise series that illustrate how a country can institutionalise exercising and progressively evolve realism, scope, and stakeholder participation over time. Two examples frequently referenced in European contexts are the *KSÖ Cybersecurity Planspiele (Austria)*, which are characterised as multi-stakeholder crisis simulations with increasingly hybrid and technically supported elements, and *PANOPTES (Greece)*, which is typically described as a national cyber defence exercise series using controlled simulation environments and repeated iterations.

Sector-led, assurance-oriented and regulatory-linked approaches. Some exercise formats are designed not only for training, but for assurance - demonstrating that response and recovery capabilities work under realistic adversary assumptions and within strict governance constraints. A prominent reference is *TIBER-EU* in the financial sector, which uses threat intelligence to guide controlled red teaming and emphasises strong governance and evidence-oriented outputs. Related approaches exist under the umbrella of *Threat-Led Penetration Testing (TLPT)* programmes, whose sector implementations vary but are generally used to validate defensive capability and governance readiness through realistic threat-driven testing.

Methodology-driven frameworks and programme templates. In addition to named exercise series, national programmes often draw on methodological reference frameworks that provide

structured artefacts, defined roles, evaluation logic, and a disciplined improvement cycle. Widely used examples include the U.S. *HSEEP* framework, which emphasises programme discipline, standard exercise documentation, evaluation planning, and improvement cycles. Complementary to this, many governance models rely on *PDCA-based* logic (Plan–Do–Check–Act) to embed exercises into continuous improvement loops that connect exercise outcomes to corrective actions and re-testing. Finally, ISO-aligned continuity and exercise guidance is frequently used as an umbrella reference for programme thinking, structured planning discipline, and systematic evaluation and improvement.

Skills development and pipeline formats. A final category consists of formats primarily used for capability building and talent development. These include national or sectoral *CTF ecosystems* and cyber challenges used for engagement and skills development, as well as *challenge-based training platforms* that offer repeatable hands-on skill-building with measurable completion. In many settings, *cyber range-based training programmes* serve as a progression layer from basic labs toward more advanced simulations. While these formats are valuable for building workforce capacity, they are often best positioned as feeder layers in national programmes rather than as primary assurance instruments for organizational or sector readiness.

9 References

Abazi, B. (2022). Establishing the national cybersecurity (resilience) ecosystem. IFAC-PapersOnLine, 55(39), 42–47.

AIT Austrian Institute of Technology. (2021). Planspiel KSÖ 2021. https://www.ait.ac.at/themen/cyber-security/projects/planspiel-ksoe-2021

Austrian Federal Government. (2021). Artificial Intelligence Mission Austria 2030 (AIM AT 2030): Strategy of the Austrian Federal Government for artificial intelligence. https://www.digitalaustria.gv.at/dam/jcr%3A44ad1b93-6358-42b8-9f651e74c9a39e7f/KI%20Strategie_AIM_AT_2030_UAbf.pdf

Austrian Federal Government. (2024). Artificial Intelligence Mission Austria 2030: Implementation plan 2024. https://www.digitalaustria.gv.at/eng/strategy/strategy-AI-AIM-AT-2030.html

Bundesamt für Sicherheit in der Informationstechnik. (2021). Kapitel 8: Notfälle üben. BSI IT-Grundschutz / Notfallmanagement. https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/IT-Grundschutz/Zertifizierte-Informationssicherheit/IT-Grundschutzschulung/Online-Kurs-Notfallmanagement/8_NotfaelleUeben/nfm08_01.html

Bundesamt für Sicherheit in der Informationstechnik. (2026). BSI-Standard 200-4: Business Continuity Management – Hilfsmittel und weiterführende Aspekte zur Bewältigung. https://www.bsi.bund.de/

Bundeskanzleramt Österreich. (2024). Österreichische Sicherheitsstrategie 2024. Republic of Austria.

Bundesministerium für Inneres. (2021). Österreichische Strategie für Cybersicherheit 2021 (ÖSCS 2021). Republic of Austria. https://www.bmi.gv.at/504/files/Cyberstrategie2021.pdf

Bundesministerium für Inneres. (2025). Maßnahmenkatalog der Österreichischen Strategie für Cybersicherheit 2021: Fortschrittsbericht 1/2025. Republic of Austria. https://www.bmi.gv.at/504/oesterreichische-strategie-fuer-cybersicherheit.html

Bundesministerium für Inneres. (2026a). Bundesamt für Cyber-Sicherheit. Republic of Austria. https://www.bmi.gv.at/magazin/2026_03_04/34_cyber_sicherheit.html

Bundesministerium für Inneres. (2026b). Nationale Strukturen der Cybersicherheit. Republic of Austria. https://www.bmi.gv.at/504/nationale-strukturen.html

Bundesrechenzentrum. (2017, November 13). Sicherheit im Netz – BRZ nimmt am KSÖ-Cybersecurity-Planspiel 2017 teil. https://www.brz.gv.at/was-wir-tun/sicherheit-und-qualitaet/sicherheit-im-netz.html

Bányász, P., Bottyán, S., Dobó, J., Dub, M., Gyaraki, R., Hankó, V., Katona, G., Kiss, A., Magas, B., & Nimsz, V. (2023). Biztonságorientáció és -tudatosítás nagyvállalati környezetben. In T. Bonnyai, A. Kiss, & A. Tóth (Eds.), Nagyvállalati biztonságtudatosság és nyílt forrású információszerzés: Kiber- és információbiztonsági tanulmányok (pp. 8–80). Ludovika Egyetemi Kiadó.

Bányász, P., Kiss, A., Magyar, S., & Kiss, D. (2024). Empirical analysis of the cyberattacks of the Russian–Ukrainian war. Információs Társadalom, 24(2), 9–32. https://doi.org/10.22503/inftars.XXIV.2024.2.1

Bányász, P., Szádeczky, T., & Váczi, K. B. (2024). The relationship between generative artificial intelligence and cybersecurity. In Proceedings of the Central and Eastern European eDem and eGov Days 2024 (pp. 209–215). Association for Computing Machinery. https://doi.org/10.1145/3670243.3670781

CERT.at. (2017). KSÖ Cybersecurity Planspiel 2017. https://www.cert.at/de/berichte/2017/2017-chap03-4

Council of the European Union. (2025). Council Recommendation of 6 June 2025 on an EU blueprint for cyber crisis management (C/2025/3445). Official Journal of the European Union, C/2025/3445. http://data.europa.eu/eli/C/2025/3445/oj

Dewar, R. S. (2018). Cybersecurity and cyberdefence exercises. Center for Security Studies, ETH Zurich.

Direktion Staatsschutz und Nachrichtendienst. (2026). Verfassungsschutzbericht 2025. Bundesministerium für Inneres. https://www.dsn.gv.at/501/files/vsb/verfassungsschutzbericht_vsb_2025_bf.pdf

European Central Bank. (2024). TIBER-EU framework: How to implement the European framework for threat intelligence-based ethical red teaming. https://www.ecb.europa.eu/pub/pdf/other/ecb.tiber_eu_framework.en.pdf

European Central Bank. (2025a). TIBER-EU framework. European Central Bank. https://www.ecb.europa.eu/pub/pdf/other/ecb.tiber_eu_framework_2025~b32eff9a10.en.pdf

European Central Bank. (2025b, November). TIBER-EU SSM implementation guide: How to implement the TIBER-EU framework for the DORA TLPT of significant institutions. https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.supervisory_guide202511.en.pdf

European Central Bank. (n.d.). What is TIBER-EU? https://www.ecb.europa.eu/paym/cyber-resilience/tiber-eu/html/index.en.html

European Commission. (2025a). AI Continent Action Plan. European Commission.

European Commission. (2025b). Commission Delegated Regulation (EU) 2025/1190 of 13 February 2025 supplementing Regulation (EU) 2022/2554 with regard to regulatory technical standards specifying the criteria for identifying financial entities required to perform threat-led penetration testing and the requirements governing TLPT. Official Journal of the European Union, L 2025/1190. https://eur-lex.europa.eu/eli/reg_del/2025/1190/oj

European Commission. (2025c). Proposal for a Council Recommendation for an EU blueprint on cybersecurity crisis management (COM(2025) 66 final).

European Commission. (n.d.). NIS Cooperation Group. Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/nis-cooperation-group

European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, L 119, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

European Parliament and Council of the European Union. (2019). Regulation (EU) 2019/881 of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and

communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act). Official Journal of the European Union, L 151, 15–69. https://eur-lex.europa.eu/eli/reg/2019/881/oj

European Parliament and Council of the European Union. (2022a). Directive (EU) 2022/2555 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 (NIS2 Directive). Official Journal of the European Union, L 333, 80–152. https://eur-lex.europa.eu/eli/dir/2022/2555/oj

European Parliament and Council of the European Union. (2022b). Directive (EU) 2022/2557 of 14 December 2022 on the resilience of critical entities and repealing Council Directive 2008/114/EC (Critical Entities Resilience Directive). Official Journal of the European Union, L 333, 164–198. https://eur-lex.europa.eu/eli/dir/2022/2557/oj

European Parliament and Council of the European Union. (2022c). Regulation (EU) 2022/2554 of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, (EU) No 909/2014 and (EU) 2016/1011 (Digital Operational Resilience Act). Official Journal of the European Union, L 333, 1–79. https://eur-lex.europa.eu/eli/reg/2022/2554/oj

European Parliament and Council of the European Union. (2024a). Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence 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, L 2024/1689. http://data.europa.eu/eli/reg/2024/1689/oj

European Parliament and Council of the European Union. (2024b). Regulation (EU) 2024/2847 of 23 October 2024 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, L 2024/2847. https://eur-lex.europa.eu/eli/reg/2024/2847/oj

European Parliament and Council of the European Union. (2025). Regulation (EU) 2025/38 of 19 December 2024 laying down measures to strengthen solidarity and capacities in the Union to detect, prepare for and respond to cyber threats and incidents and amending Regulation (EU) 2021/694 (Cyber Solidarity Act). Official Journal of the European Union, L 2025/38. http://data.europa.eu/eli/reg/2025/38/oj

European Security and Defence College. (n.d.). EAB.Cyber – Cyber Education, Training, Exercise and Evaluation Platform. https://www.esdc.europa.eu/configurations/eabcyber_en

European Union Agency for Cybersecurity. (2024). Cyber Europe 2024 – After Action Report. ENISA.

European Union Agency for Cybersecurity. (2026). The ENISA cybersecurity exercise methodology: End-to-end guide on how to plan, run and evaluate an exercise (Version 1.0). https://www.enisa.europa.eu/publications/the-enisa-cybersecurity-exercise-methodology

European Union Agency for Cybersecurity. (n.d.-a). CSIRTs Network. ENISA. https://www.enisa.europa.eu/topics/eu-incident-response-and-cyber-crisis-management/csirts-network

European Union Agency for Cybersecurity. (n.d.-b). EU-CyCLONe. ENISA. https://www.enisa.europa.eu/topics/eu-incident-response-and-cyber-crisis-management/eu-cyclone

Europol. (2022). Facing reality? Law enforcement and the challenge of deepfakes: An observatory report from the Europol Innovation Lab. Publications Office of the European Union. https://doi.org/10.2813/158794

Federal Emergency Management Agency. (2020). Homeland Security Exercise and Evaluation Program (HSEEP). U.S. Department of Homeland Security. https://preptoolkit.fema.gov/web/hseep-resources/policy-and-guidance

Finanzmarktaufsicht Österreich. (2025, January 16). Neuer EU-Rahmen macht Finanzsystem fit gegen Cyberangriffe und IT-Ausfälle – DORA-Start im Fokus der FMA. https://www.fma.gv.at/quantensprung-fuer-ikt-sicherheit-dora-anwendbar/

Kenedli, T., & Magyar, S. (Eds.). (2025). A mesterséges intelligencia hatásainak specifikus vizsgálata. Katonai Nemzetbiztonsági Szolgálat. https://knbsz.gov.hu/kiadvanyok

Kompetenzzentrum Sicheres Österreich. (n.d.). Planspiele und Simulationsübungen. https://kompetenzzentrum-sicheres-oesterreich.at/projekte/planspiel/

Kovács, L. (2018). A kibertér védelme. Dialóg Campus Kiadó. https://www.uni-nke.hu/document/uni-nke-hu/Kov%C3%A1cs%20L%C3%A1szl%C3%B3.pdf

Kovács, L. (2021). Offenzív kiberműveletek II.: Kibererők és képességeik. Hadmérnök, 16(3), 119–137. https://doi.org/10.32567/hm.2021.3.7

Kovács, L. (2023a). Hadviselés a 21. században: Kiberműveletek. Ludovika Egyetemi Kiadó. https://doi.org/10.36250/01057_00

Kovács, L. (2023b). Nyomásgyakorlás a kritikus információs infrastruktúrák támadásán keresztül. In C. Krasznay (Ed.), Taktikák és stratégiák a kiberhadviselésben (pp. 151–168). Ludovika Egyetemi Kiadó.

Kovács, L., & Krasznay, C. (2024). Digitális Mohács 3.0. Hadtudomány, 34(3), 40–55. https://doi.org/10.17047/HADTUD.2024.34.3.40

Kozák, K. (2026). Tóth András – A digitális átalakulás szabályozása az Európai Unióban. In Medias Res, 15(1), 195–204.

Krasznay, C. (2024a). Hacktivists, proxy groups, cyber volunteers: The future of non-state actors' involvement in military cyber operations. Academic and Applied Research in Military and Public Management Science, 23(3), 107–124. https://doi.org/10.32565/aarms.2024.3.6

Krasznay, C. (2024b). The role of civilian cybersecurity companies in military cyber operations. Land Forces Academy Review, 29(1), 1–10. https://doi.org/10.2478/raft-2024-0001

Krasznay, C. (2025). A NIS2 irányelv kihívásai és gyakorlati alkalmazása. Védelem Tudomány, 9 (Különszám), 1–5. https://ojs3.mtak.hu/index.php/vedelemtudomany/article/view/18114

Krasznay, C., & Gyebnár, G. (2021). Possibilities and limitations of cyber threat intelligence in energy systems. In 2021 13th International Conference on Cyber Conflict (CyCon) (pp. 171–188). IEEE. https://doi.org/10.23919/CyCon51939.2021.9468289

Leitner, M., Páhi, T., & Skopik, F. (2017). Situational awareness for strategic decision making on a national level. In F. Skopik (Ed.), Collaborative cyber threat intelligence (pp. 225–276). CRC Press.

Leitner, M., Páhi, T., & Skopik, F. (2024). Operational cyber incident coordination revisited: Providing cyber situational awareness across organizations and countries. Information Security Journal: A Global Perspective, 33(5), 486–507.

Maathuis, C., & Cools, K. (2025). The role of AI in military cyber security: Data insights and evaluation methods. Procedia Computer Science, 254, 191–200.

Magyar, S., Bányász, P., Bányász-Váczi, K. B., & Pál, A. (2025). Magyarország kibervédelmi stratégiáinak összehasonlítása és fejlődési íve 2013 és 2025 között. Szakmai Szemle, 23(1), 127–144. https://ojs.knbsz.gov.hu/index.php/szakmai/article/view/228

Magyar, S., & Péter, K. F. (2025). A mesterséges intelligencia hatásainak specifikus vizsgálata című tudományos–szakmai konferencia és könyvbemutató. Szakmai Szemle, 23(3–4). https://www.knbsz.gov.hu/ac57d5f9-b367-4a2a-a604-5e7309961a6d

Marsi, T. (2025, December 17). 2026 a kiberreziliencia éve. ICT Global Magazin. https://ictglobal.hu/print/kiberbiztonsag-mi-2026-marsi-tamas/

Molnár, D. (2022). Great power cyber diplomacy: China on the international cyber platform accompanied by the United States and Russia. Hadtudomány, 32(E-szám), 227–241. https://doi.org/10.17047/Hadtud.2022.32.E.227

Nagy, V. (2025). A generatív mesterséges intelligencia felhasználásának területei a nemzetbiztonsági elemző-értékelő munkában, különös tekintettel a döntéstámogatásra. In T. Kenedli & S. Magyar (Eds.), *A mesterséges intelligencia hatásainak specifikus vizsgálata* (pp. 348–369). Katonai Nemzetbiztonsági Szolgálat. https://knbsz.gov.hu/kiadvanyok

NATO Allied Command Transformation. (n.d.-a). Cyber Coalition. https://www.act.nato.int/activities/cyber-coalition/

NATO Allied Command Transformation. (n.d.-b). Our work. https://www.act.nato.int/our-work/

NATO Cooperative Cyber Defence Centre of Excellence. (n.d.-a). Crossed Swords. https://ccdcoe.org/crossed-swords/

NATO Cooperative Cyber Defence Centre of Excellence. (n.d.-b). The NATO Cooperative Cyber Defence Centre of Excellence. https://ccdcoe.org/

North Atlantic Treaty Organization. (2024, July 30). Cyber defence. https://www.nato.int/en/what-we-do/deterrence-and-defence/cyber-defence

Olakojo, K. A. (2024). Strengthening cross-border cybersecurity resilience through federated threat intelligence sharing, real-time incident correlation, and coordinated governance mechanisms. International Journal of Computer Applications Technology and Research, 13(12), 138–152. https://doi.org/10.7753/IJCATR1312.1011

Páhi, T., & Skopik, F. (2016). A public-private-partnership model for national cyber situational awareness. International Journal on Cyber Situational Awareness, 1, 1–23.

Páhi, T., & Tóth, A. (2024). The potential of cybersecurity exercises in critical infrastructures. American Journal of Research Education and Development, 2024(3), 13–48.

Páhi, T., Leitner, M., & Skopik, F. (2017). Analysis and assessment of situational awareness models for national cyber security centers. In P. Mori, S. Furnell, & O. Camp (Eds.), Proceedings of the 3rd International Conference on Information Systems Security and Privacy (ICISSP) (pp. 334–345). SciTePress.

Páhi, T., Skopik, F., Kieseberg, P., & Leitner, M. (2018). Erhebung von Informations- und Datenquellen für Cyber-Lagebilder. In F. Skopik, T. Páhi, & M. Leitner (Eds.), Cyber situational awareness in public-private-partnerships: Organisationsübergreifende Cyber-Sicherheitsvorfälle effektiv bewältigen (pp. 191–236). Springer Vieweg.

Páhi, T., Tóth, A., & Tjoa, S. (2025). Evaluating national cyber exercise programs. In 2025 IEEE 19th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 553–558). IEEE.

Radanliev, P. (2025). Cyber diplomacy: Defining the opportunities for cybersecurity and risks from artificial intelligence, IoT, blockchains, and quantum computing. Journal of Cyber Security Technology, 9(1), 28–78.

Republic of Austria. (2024). DORA-Vollzugsgesetz (DORA-VG), BGBl. I Nr. 112/2024. Rechtsinformationssystem des Bundes. https://www.ris.bka.gv.at/eli/bgbl/I/2024/112

Republic of Austria. (2025a). Bundesgesetz zur Gewährleistung eines hohen Cybersicherheitsniveaus von Netz- und Informationssystemen (Netz- und Informationssystemsicherheitsgesetz 2026 – NISG 2026), BGBl. I Nr. 94/2025. Rechtsinformationssystem des Bundes. https://ris.bka.gv.at/eli/bgbl/I/2025/94

Republic of Austria. (2025b). Bundesgesetz zur Sicherstellung eines hohen Resilienzniveaus von kritischen Einrichtungen (Resilienz kritischer Einrichtungen-Gesetz – RKEG), BGBl. I Nr. 60/2025. Rechtsinformationssystem des Bundes. https://ris.bka.gv.at/eli/bgbl/I/2025/60

Republic of Austria. (2026). Resilienz kritischer Einrichtungen-Verordnung (RKEV), BGBl. II Nr. 91/2026. Rechtsinformationssystem des Bundes. https://www.ris.bka.gv.at/eli/bgbl/ii/2026/91

Ristvej, J., Tonhauser, M., Chovanec, D., Kubás, J., Kollár, B., & Zamiar, Z. (2025). Cyber resilience conceptual model for European Union NIS2 standards implementation in Slovakia. Scientific Reports, 15, Article 26902. https://doi.org/10.1038/s41598-025-12829-3

Rundfunk und Telekom Regulierungs-GmbH. (2024). Service Desk for Artificial Intelligence. RTR. https://www.rtr.at/rtr/service/ki-servicestelle/KI-Servicestelle.en.html

Rundfunk und Telekom Regulierungs-GmbH. (2025). Kommunikationsbericht 2024: Servicestelle für Künstliche Intelligenz. https://www.rtr.at/rtr/publikationen/Kommunikationsbericht/RTR_Kommunikationsbericht_24_WebKorr.pdf

Saharinen, K. (2026, February). JYVSECTEC and the 2026 National Cyber Exercise for Municipalities and Critical Infrastructure Operators (KYHA) [Presentation]. JAMK University of Applied Sciences.

Tabansky, L., & Lichterman, E. (2025). PROGRESS: The sectoral approach to cyber resilience. International Journal of Information Security, 24, Article 18. https://doi.org/10.1007/s10207-024-00910-3

Turaga, V. S. P., Gopali, S., Kieseberg, P., Páhi, T., Tjoa, S., Jones, K. S., & Siami Namin, A. (2026). ACT: A conceptual framework for understanding and controlling trustworthiness in controllable AI. IEEE Access, 14, 84016–84066.

Turaga, V. S. P., Páhi, T., Tjoa, S., Siami-Namini, S., & Siami Namin, A. (2025). CO2 (Co-Compliance Officer): An LLM-based ontology-driven methodology for generating knowledge graphs and AI compliance checking. IEEE Access, 13, 210576–210606.

Tóth, A. (2022). A digitális állam információbiztonsági kihívásai. Ludovika Egyetemi Kiadó. https://doi.org/10.36250/01048_00

Tóth, A. (2024). A digitális átalakulás szabályozása az Európai Unióban. ORAC Kiadó.

Ziring, N. (2022). National cyber resilience and roles for public and private sector stakeholders. In International Conference on Critical Infrastructure Protection (pp. 3–46). Springer Nature Switzerland.