Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AI-Based Scenario Generation in Cyber Training Ranges: A Systematic Review

Purvi Mehta

purvim@students.federation.edu.au

Federation University

Paul Black

Federation University

Gour Karmakar

Federation University

Paul Pang

Federation University

Research Article

Keywords: Cyber Training Range, Scenario Generation, Artificial Intelligence, Large Language Models, Retrieval-Augmented Generation, Cyber Threat Intelligence

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11421747/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

AI-Based Scenario Generation in Cyber Training Ranges: A Systematic Review

Purvi Mehta[1*], Paul Black[1], Gour Karmakar[1], Paul Pang[1]

[1]Internet Commerce Security Lab, Federation University, Australia.

*Corresponding author(s). E-mail(s): purvim@students.federation.edu.au;
Contributing authors: p.black@federation.edu.au;
gour.karmakar@federation.edu.au; p.pang@federation.edu.au;

Abstract

Developing realistic scenarios for Cyber Training Ranges (CTRs) requires current threat information, suitable infrastructure, attack events, learning objectives, and consideration of trainee capability. Established scenario-generation approaches use manual design, simulation, templates, rules, attack graphs, and model-driven engineering, often requiring substantial expert effort to develop and update. Recent research has investigated Artificial Intelligence (AI) for scenario generation and training, including Machine Learning (ML), Reinforcement Learning (RL), and Large Language Models (LLMs). However, existing surveys provide fragmented coverage of AI across the scenario-generation lifecycle and have not systematically examined AI-specific threat intelligence, such as MITRE ATLAS, as a dedicated source for CTR scenarios. This paper presents a Systematic Literature Review (SLR) of embedding AI into CTR scenario generation, tracing the progression from non-AI approaches to ML, LLM-based, and agentic AI methods. It critically compares their advantages and limitations and examines Retrieval-Augmented Generation (RAG), agentic approaches, and AI threat intelligence-based attack scenarios. The literature indicates that AI can support scenario creation, automation, adaptation, and deployment, while important challenges remain in threat-data quality and currency, hallucinations, technical validation, reproducibility, interoperability, and alignment with training objectives. The review identifies these limitations and discusses research directions for more realistic, adaptable, and technically reliable cybersecurity training scenarios.

Keywords: Cyber Training Range, Scenario Generation, Artificial Intelligence, Large Language Models, Retrieval-Augmented Generation, Cyber Threat Intelligence

1 Introduction

Cybersecurity threats continue to evolve as organisations become increasingly dependent on interconnected systems, cloud services, operational technologies, and digital infrastructure. This creates an ongoing need for cybersecurity professionals to develop practical skills in identifying, analysing, and responding to attacks. Classroom-based instruction can provide the required theoretical knowledge, but it cannot fully reproduce the technical and operational conditions encountered during a real cyber incident. Cyber Training Ranges (CTRs) address this limitation by providing controlled environments in which participants can practice offensive and defensive activities using simulated or emulated networks, systems, services, and attack scenarios (Yamin et al. 2020; Ukwandu et al. 2020; Chouliaras et al. 2021). These environments allow trainees to experience realistic cybersecurity incidents while avoiding the risks associated with conducting such activities on operational systems.

The effectiveness of a CTR depends not only on the underlying infrastructure but also on the quality of the scenarios used for training. A scenario determines the context in which an exercise takes place, the systems and entities involved, the sequence of events and attacks, the actions expected from participants, and the learning objectives to be achieved. Developing such scenarios is therefore more than a technical deployment task. Instructors must consider the threat being represented, the configuration of the training environment, the skills of the participants, and how the scenario should progress as the exercise unfolds. Existing CTRs address these requirements through different approaches, ranging from manually designed and rule-based scenarios to model-driven generation and automated deployment (Garg et al. 2025; Yamin et al. 2020). However, preparing and maintaining realistic scenarios can require considerable expertise and effort, particularly when exercises need to be updated for emerging threats or adapted for different training objectives and levels of experience.

These limitations have encouraged researchers to investigate how Artificial Intelligence (AI), such as traditional Machine Learning (ML) and Deep Learning (DL), can support scenario generation and management in CTRs. Rather than replacing the complete scenario-development process, AI can assist with particular tasks, such as analysing threat information, generating or adapting scenario content, modelling adversarial behaviour, and adjusting training activities according to the exercise context. The main limitations of ML include the limited availability of labelled training data and its dependence on historical data used during model development. More recently, Large Language Models (LLMs) have extended this capability by generating scenario narratives and other exercise content from natural-language instructions. However, an LLM relies substantially on knowledge acquired during training and may produce inaccurate, outdated, or technically inconsistent information. Retrieval-Augmented Generation (RAG) can augment an LLM by retrieving relevant external information, such as Cyber Threat Intelligence (CTI), vulnerability information, attack reports, and organisational documentation, and supplying this information as additional context for generation. RAG can improve the relevance and currency of generated scenarios, although the resulting output still depends on the quality of the retrieved information and requires appropriate validation before use in a CTR.

Several reviews have examined CTRs from different perspectives, including their architectures, technologies, taxonomies, educational applications, scenario-management capabilities, and supporting infrastructure (Davis and Magrath 2013; Yamin et al. 2020; Ukwandu et al. 2020; Chouliaras et al. 2021). More recent studies have extended this discussion to areas such as scenario description languages, CTI integration, interoperability, reference architectures, and scenario lifecycle management (Garg et al. 2025; Kampourakis et al. 2025; Lillemets et al. 2025). These studies provide an important foundation for understanding how CTRs are designed and operated, but AI-enabled scenario generation is generally considered only as one component of a broader CTR discussion or through individual technologies and stages of the scenario lifecycle. Consequently, there remains limited synthesis of how AI techniques are being incorporated across scenario design, generation, validation, deployment, adaptation, and assessment. Table 1 summarises the scope and contributions of representative CR review and survey studies and compares them with the focus of this review. Table 1 shows that the treatment of AI in existing surveys is generally limited to individual technologies or to selected stages of the scenario-generation lifecycle. Additionally, existing surveys lack a systematic focus on AI-specific threat intelligence, such as MITRE ATLAS, as a primary source for developing CTR scenarios. This review addresses this gap by systematically comparing AI techniques across the complete scenario-generation lifecycle.

Against this background, this review examines how scenario generation for CTRs has evolved from established non-AI approaches to AI-supported methods. The review has three main objectives. First, it identifies and compares the principal approaches used to generate CTR scenarios and examines their strengths and limitations. Second, it investigates how AI is being incorporated into scenario generation, including its role in scenario creation, adaptation, validation, deployment, and assessment. Particular attention is given to recent LLM-based approaches, including the use of RAG to ground generated content in external cybersecurity knowledge, and to emerging agentic approaches that extend generation towards more autonomous and knowledge-based attack generation and scenario-management workflows. Third, the review analyses the limitations reported across these approaches and identifies research challenges that must be addressed to support realistic, adaptable, scalable, and technically reliable CTR scenarios.

The major contributions of this review are as follows:

1. **Scenario-generation taxonomy:** We organise existing CTR scenario-generation methods into a structured taxonomy covering non-AI approaches and emerging AI-supported approaches, showing how scenario development has progressed from manually defined and model-driven methods towards more adaptive and automated techniques.
2. **Analysis of AI-supported scenario generation:** We examine how AI is being applied across the scenario-generation lifecycle, including scenario creation, adaptation, validation, deployment, and assessment. Particular attention is given to LLM-based generation, the use of RAG to ground LLM outputs in external cybersecurity knowledge, and emerging agentic approaches that extend automation

through knowledge-based attack generation, reasoning, planning, retrieval, tool use, and iterative validation. In addition, existing surveys have not systematically examined AI-specific threat intelligence, such as MITRE ATLAS, as a dedicated source for generating CTR scenarios (see Table 1).

3. **Comparative analysis of existing approaches:** We compare representative non-AI and AI-supported approaches in terms of their underlying methods, contributions, advantages, limitations, and reported or suggested future improvements. This comparison is used to identify where AI provides capabilities that are difficult to achieve through static or predefined scenario-generation methods.
4. **Identification of research challenges and future directions:** Based on the reviewed literature, we synthesise unresolved issues including data quality, formal and technical validation, reproducibility, interoperability, scenario portability, learning-objective alignment, adaptability, evaluation, explainability, and safety, and identify directions for the development of more reliable and intelligent CTRs.

2 Scope of the Systematic Literature Review

Existing CTR surveys have provided valuable insights into range architectures, technologies, functional components, educational applications, and scenario-management capabilities. More recent studies have also examined scenario description languages, CTI integration, reference architectures, interoperability, accessibility, and scenario lifecycle management. However, these topics are generally considered within broader reviews of CTRs, while the role of AI across the complete scenario-generation lifecycle remains fragmented. In particular, there is limited synthesis comparing conventional ML and DL techniques, adaptive approaches such as Reinforcement Learning (RL), generative AI, and LLM-based scenario generation, including RAG-enhanced and agentic architectures, across scenario design, generation, validation, deployment, adaptation, and assessment. This Systematic Literature Review (SLR) therefore focuses specifically on the evolution and application of AI-supported scenario generation in CTRs while retaining relevant non-AI approaches as a technical baseline (Ukwandu et al. 2020; Yamin et al. 2020; Chouliaras et al. 2021; Garg et al. 2025; Kampourakis et al. 2025; Lillemets et al. 2025).

2.1 Objectives of the Review

The primary objective of this SLR is to critically examine the evolution and current state of AI-supported scenario generation in CTRs. The review investigates how conventional ML and DL techniques, adaptive approaches such as RL, generative AI, and LLM-based methods contribute to different stages of the scenario-generation lifecycle. Particular attention is given to RAG-enhanced LLM applications and emerging agentic architectures that extend scenario generation towards grounded and increasingly automated workflows. The review compares these approaches with established non-AI methods, evaluates their reported strengths and limitations, and identifies research challenges affecting the development of realistic, adaptable, scalable, and technically reliable CTR scenarios.

Table 1: Comparison of Existing Cyber-Range Surveys with the Proposed Review.

Survey	Scenario Models	AI/ ML	LLM-based	RAG-enhanced	AI Threat Intel.	CTI	Formal Validation	Adaptive Training	Principal Focus / Gap
Davis and Magrath (2013) (Davis and Magrath 2013)	△	×	×	×	×	×	×	×	Foundational CR/testbed classification; predates AI-driven scenario and adaptive attack generation.
Ukwandu et al. (2020) (Ukwandu et al. 2020)	✓	△	×	×	×	△	△	△	Broad CR/testbed review; covers automation and emerging technologies but not AI-generated or adaptive attack scenarios as a distinct research stream.
Yamin et al. (2020) (Yamin et al. 2020)	✓	×	×	×	×	△	△	△	Comprehensive CR taxonomy and scenario lifecycle; intelligent scenario generation and adaptive AI attackers remain outside its principal scope.
Chouliaras et al. (2021) (Chouliaras et al. 2021)	✓	×	×	×	×	△	△	△	Reviews CR platforms for education, training, and research; provides limited treatment of AI-driven scenario or attack generation.
Garg et al. (2025) (Garg et al. 2025)	✓	✓	✓	△	×	✓	✓	△	End-to-end scenario-generation review covering CTI and AI/ML-enabled methods, including RL-based virtual attackers. Its CTI-driven scenario generation primarily uses conventional cyber threat intelligence sources, such as MITRE ATT&CK, rather than AI-specific threat intelligence frameworks such as MITRE ATLAS.
Kampourakis et al. (2025) (Kampourakis et al. 2025)	✓	△	△	×	✓	△	△	△	Reference architecture for CR components and standardisation; includes AI-assisted scenario generation and AI-driven threat intelligence, but does not systematically review AI-based scenario-generation methods.
Lillemets et al. (2025) (Lillemets et al. 2025)	✓	✓	✓	×	×	△	×	△	Updated CR taxonomy incorporating AI-based and autonomous agents, AI-driven scenario adaptation, and emerging LLM use. However, it does not examine AI-specific threat intelligence frameworks such as MITRE ATLAS as a dedicated source for scenario or attack generation.
Proposed Review	✓	✓	✓	✓	✓	✓	✓	✓	**Unifies non-AI, AI, LLM/RAG, and AI threat intelligence-based, AI-generated or adaptive attack approaches across the CTR scenario lifecycle, exposing validation, adaptability, and deployment gaps.**

Legend: ✓ = substantial coverage; △ = partial or indirect coverage; × = not a principal focus.

Within this broader objective, the review also examines two complementary directions for threat-informed scenario generation. First, it considers how MITRE Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK) can support the representation and generation of attack scenarios targeting traditional Information and Communication Technology (ICT) systems and networks. Second, it considers how MITRE ATLAS can support scenario generation for adversarial threats targeting AI and ML systems, including LLM-based applications (MITRE 2026; Sachpelidis-Brozos et al. 2026). These two directions provide a basis for identifying future research opportunities in CTR scenario generation for both conventional ICT environments and emerging AI-enabled systems.

2.2 Study Selection Process

The SLR follows the PRISMA 2020 guidelines to provide a transparent and systematic process for identifying, screening, assessing, and selecting relevant studies (Page et al. 2021). Predefined inclusion and exclusion criteria were applied throughout the selection process to ensure that the studies aligned with the scope and objectives of this review. The complete study-selection process is illustrated in Fig. 1.

The review examines the evolution of scenario-generation approaches in CTRs, with particular emphasis on the integration of AI across the scenario lifecycle. Non-AI approaches are also included where they provide an important technical baseline for understanding the progression towards automated and AI-supported scenario generation.

Inclusion and Exclusion Criteria: Studies were assessed using predefined inclusion and exclusion criteria designed to retain primary studies that contribute directly to CTR scenario generation or its supporting lifecycle. Eligible studies included technical methods, frameworks, architectures, implementations, or evaluated approaches addressing scenario modelling, generation, automation, orchestration, validation, deployment, adaptive training, or AI integration. Relevant non-AI studies were retained where they provided a technical baseline for comparison with AI-supported approaches. Studies were excluded when they were unrelated to CTRs, lacked a substantive technical contribution, focused on general cybersecurity or AI without a clear CTR application, or provided insufficient methodological or technical detail for analysis. The complete criteria are summarised in Table 2.

1. **Identification:** A systematic search was conducted across six scientific databases: IEEE Xplore, ACM Digital Library, Scopus, Web of Science, ScienceDirect, and SpringerLink.

   Search Strategy and Keywords: Search terms were organised around four concept groups: (i) cyber training environments, (ii) scenario generation and management, (iii) technical approaches to scenario generation, and (iv) AI-related approaches to scenario generation. Synonymous and closely related terms within each concept group were combined using the Boolean operator `OR`, while the relevant concept groups were combined using `AND`. The search syntax was adapted to each database's requirements while retaining the same underlying search

Table 2: Inclusion and Exclusion Criteria for Study Selection

Inclusion Criteria	Exclusion Criteria
Studies directly related to CTRs and their scenario or training lifecycle.	Studies unrelated to CTRs or cybersecurity training environments.
Primary studies presenting a technical method, framework, architecture, implementation, experiment, or evaluated approach.	Surveys, reviews, bibliometric studies, editorials, opinion papers, or other non-primary studies.
Studies integrating AI or related intelligent techniques into CTRs.	General AI or cybersecurity studies without a clear CTR application.
Non-AI studies providing a technical baseline for scenario modelling, generation, automation, orchestration, validation, or deployment.	Studies that only mention AI or CTRs as future work without a substantive technical contribution.
Studies providing sufficient technical or methodological information for assessing their contribution.	Studies lacking sufficient technical or methodological information for reliable assessment.
Studies published in English and providing sufficient information for eligibility assessment.	Studies not published in English.

concepts. The cyber training environment and scenario-generation concept groups were combined as follows:

("cyber range" OR "cyber ranges" OR "cyber training range" OR "cyber security range" OR "cybersecurity range" OR "security testbed" OR "security testbeds" OR "cyber testbed" OR "cyber testbeds" OR "cyber exercise" OR "cyber security exercise" OR "cybersecurity exercise") AND ("scenario generation" OR "scenario design" OR "scenario modelling" OR "scenario modeling" OR "scenario creation" OR "scenario specification" OR "scenario automation" OR "scenario instantiation" OR "scenario deployment" OR "scenario orchestration" OR "scenario adaptation" OR "scenario execution" OR "scenario validation" OR "scenario testing" OR "scenario randomization" OR "scenario randomisation" OR "scenario configuration" OR "scenario modelling and execution" OR "exercise scenario generation" OR "exercise scenario design" OR "exercise scenario modelling" OR "exercise scenario modeling" OR "exercise content generation" OR "training scenario generation" OR "training scenario design" OR "topology generation")

To capture established scenario-generation approaches that do not necessarily use AI terminology, the following technical terms were also considered:

("model-driven" OR "model driven" OR "rule-based" OR "rule based" OR "template-based" OR "template based" OR "domain-specific language" OR "domain specific language" OR "scenario definition language" OR "declarative specification" OR "formal verification" OR "ontology-based" OR "ontology based" OR "platform-independent" OR "platform independent" OR "platform-agnostic" OR "platform agnostic" OR "infrastructure-as-code" OR "infrastructure as code")

To capture studies using artificial intelligence in CTR scenario generation, a separate AI-related concept group was defined as follows:

("artificial intelligence" OR "AI" OR "machine learning" OR "deep learning" OR "reinforcement learning" OR "generative AI" OR "generative artificial intelligence" OR "large language model" OR "large language models" OR "LLM" OR "LLMs" OR "retrieval augmented generation" OR "retrieval-augmented generation" OR "RAG" OR "agentic AI" OR "AI agent" OR "AI agents" OR "agentic RAG" OR "multi-agent system" OR "multi-agent systems" OR "multi-agent" OR "AI-driven threat intelligence" OR "AI threat intelligence" OR "threat forecasting")

The primary search combined the cyber training environment terms with the scenario-generation terms using the Boolean operator AND. This broad search was used to identify relevant studies without requiring a specific technical approach. Two supplementary targeted search structures were also used: one incorporating established technical terms and another incorporating AI-related terms. These targeted searches complemented the primary search by identifying studies that described scenario-generation methods using more specific technical terminology. The targeted searches followed two additional Boolean structures:

Traditional technical search:
(cyber-range terms) AND *(scenario-generation terms)* AND *(traditional technical terms)*

AI-related search:
(cyber-range terms) AND *(scenario-generation terms)* AND *(AI-related terms)*

Studies published between 2011 and 2026 were considered. Fig. 1 shows the database searches identified 1,616 records: 105 from IEEE Xplore, 220 from ACM Digital Library, 228 from Scopus, 130 from Web of Science, 403 from ScienceDirect, and 530 from SpringerLink. After removing 287 duplicate records, 1,329 unique records remained for title and abstract screening.

2. **Screening:** The titles and abstracts of the 1,329 records were screened using the predefined inclusion and exclusion criteria. Records clearly unrelated to CTRs, scenario generation, cybersecurity training, or relevant AI and automation approaches were excluded. Where relevance could not be determined confidently from the title and abstract, the record was retained for further assessment. This screening stage excluded 1,251 records, leaving 78 reports for full-text assessment.
3. **Eligibility:** Following title and abstract screening, 78 reports were retained for full-text assessment. All 78 reports were successfully retrieved and assessed against the predefined eligibility criteria, with particular attention to their contribution to CTR scenario generation, modelling, automation, deployment, adaptive training, intelligent management, or AI integration. During the full-text assessment, one duplicate or overlapping publication was identified, resulting in 77 unique reports. Of these, 24 reports were excluded because they did not satisfy the eligibility criteria, leaving 53 studies eligible for inclusion through the systematic database-search pathway.
4. **Supplementary Literature Identification:** Following the database-based selection process, supplementary literature identification was conducted to capture relevant studies not identified through the database search pathway. The same eligibility criteria were applied when assessing these studies for relevance to the

scope of the review. Four additional studies addressing relevant scenario generation and automation approaches were identified through supplementary searches. In addition, one scholarly review examining the MITRE ATLAS framework and adversarial threats to AI systems was retained because of its relevance to AI-oriented threat modelling and the development of realistic cyber-training scenarios. These five supplementary studies were combined with the 53 studies retained through the database-search pathway, resulting in a final review corpus of 58 studies.

5. **Inclusion:** The systematic database-search pathway yielded 53 eligible studies. Supplementary literature identification yielded an additional 5 studies, resulting in a final systematic review corpus of 58 studies. These 58 studies form the principal evidence base for the comparative analysis presented in this review. Additional literature is cited throughout the paper to establish background concepts, describe existing surveys and frameworks, support the methodology, and contextualise the findings; these references are not counted in the systematic-review corpus.

2.3 Analysis of Included Studies and Research Trends

The studies included in the systematic review were analysed to explore the development and technological progress of scenario-generation methods in CTRs. To provide a broader historical view of this technological evolution, the trend analysis presented in Fig. 2 also considers relevant technical studies identified in the broader literature and cited in this review. These studies were classified by publication year and principal technological approach as non-AI, traditional AI, or LLM-based. This broader analysis is used specifically to illustrate the evolution of scenario-generation technologies and does not alter the systematic-review corpus reported in the PRISMA analysis.

Fig. 2 shows that AI-based CTR scenario models, particularly those using LLMs, have become increasingly prevalent in recent years, especially since 2024.

3 Cyber Training Ranges Overview

CTRs, also referred to as Cyber Ranges (CRs), are controlled environments designed to replicate realistic Information Technology (IT), Operational Technology (OT), and Cyber–Physical System (CPS) infrastructures for cybersecurity training, experimentation, validation, and research. They provide safe and isolated environments in which students, cybersecurity professionals, researchers, and organizations can practice offensive and defensive techniques, evaluate security controls, and develop incident-response capabilities without exposing operational systems to unnecessary risk. Early cyber ranges were primarily designed for Computer Network Operations (CNO) training and security experimentation through simulation and emulation techniques (Davis and Magrath 2013). As cyber threats and supporting technologies evolved, modern cyber ranges expanded to support red-team and blue-team exercises, vulnerability assessment, cyber capability evaluation, security research, and collaborative training across academic, government, military, and industrial environments (Priyadarshini 2018; Ukwandu et al. 2020; Zhang et al. 2023; Stamatopoulos et al. 2024).

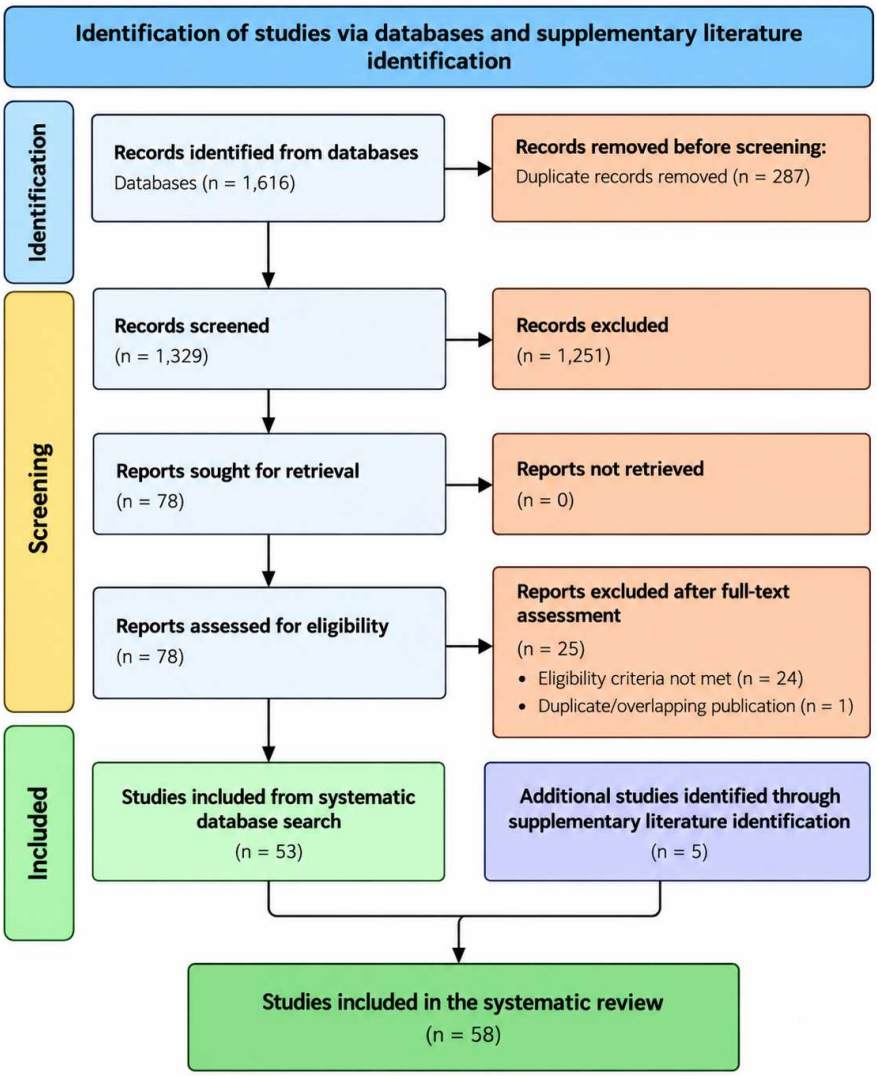


Fig. 1: PRISMA-based study-selection process used in this systematic literature review, adapted from the PRISMA 2020 framework (Page et al. 2021).

3.1 Background

CTRs provide controlled environments in which cybersecurity professionals can develop and evaluate practical skills without exposing operational systems to unnecessary risk. Through virtualised or emulated infrastructure, a CTR can reproduce networks, hosts, services, vulnerabilities, and attack conditions required for cybersecurity exercises. Automation further supports the provisioning and configuration of training environments, execution of exercise events, monitoring of participant activities, and restoration

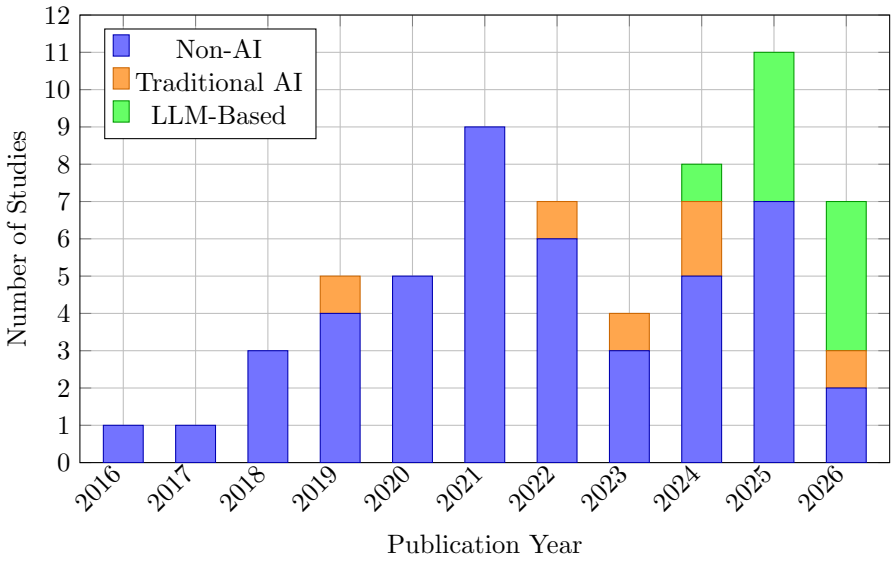


Fig. 2: Evolution of non-AI, traditional AI, and LLM-based approaches in relevant cyber-range scenario-generation literature.

of the environment for subsequent exercises (Pham et al. 2016; Gustafsson and Almroth 2020).

Automation is particularly important as CTR scenarios increase in scale and complexity. Automated provisioning and orchestration can reduce the effort required to prepare and execute exercises, support repeatable environment configuration, and facilitate the validation and restoration of training environments. These capabilities provide an important technical foundation for scenario generation because a generated scenario must ultimately be translated into an environment that can be configured, deployed, executed, and monitored within the CTR (Gustafsson and Almroth 2020).

CTR scenarios may contain both static and dynamic elements. Static elements define the initial environment, such as network topology, hosts, services, users, and vulnerabilities, whereas dynamic elements introduce changes during scenario execution. These may include simulated attacks, traffic generation, system events, or actions performed by participating teams. Consequently, a dynamic scenario may evolve according to events and participant actions rather than following a completely predetermined sequence (Yamin et al. 2020).

Modern CTRs combine technical infrastructure, scenario-management capabilities, user management, monitoring, and learning functions to support realistic cybersecurity exercises. Although implementations differ across platforms, as shown in Fig. 3, the literature identifies several recurring components that enable the creation, execution, management, and evaluation of training scenarios (Kampourakis et al. 2025; Lillemets et al. 2025).

Range Learning Management System (RLMS): The learning-management component supports the educational functions of a CTR, including the organisation of

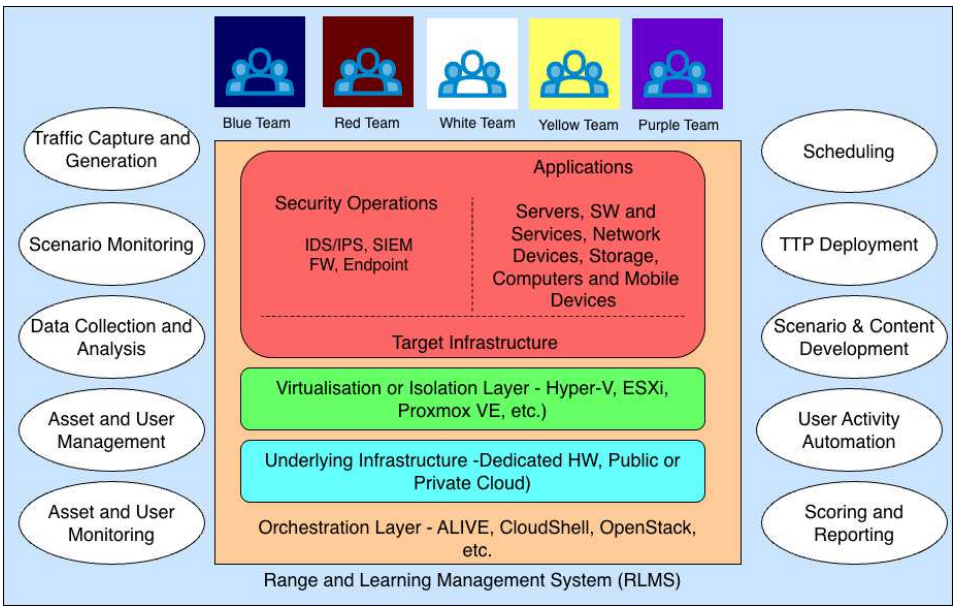


Fig. 3: Components of a Cyber Training Range, adapted from *The Cyber Range: A Guide* National Institute of Standards and Technology (2023).

training activities, management of participants, delivery of exercise information, and evaluation of learning outcomes. This educational layer helps distinguish a training-oriented cyber range from a conventional cybersecurity testbed (Lillemets et al. 2025).

Scenario Development and Management: Scenario-management functions support the creation, configuration, execution, and modification of cybersecurity exercises. A scenario may specify the technical environment, participant roles, vulnerabilities, attack events, exercise objectives, and expected outcomes. Increasing automation in this component has enabled scenarios to progress from manually prepared exercises towards model-driven, automated, and more recently AI-supported generation (Russo et al. 2018; Yamin et al. 2024).

User, Asset, Scoring, and Reporting Management: CTRs require mechanisms for managing participants and the computing assets used during an exercise. Monitoring and scoring functions capture participant actions and exercise events, allowing instructors to evaluate performance and conduct post-exercise analysis. These functions are particularly important for connecting scenario execution with assessment and subsequent refinement of training activities (Lillemets et al. 2025).

Core Infrastructure and Orchestration: The underlying infrastructure provides the computing, networking, storage, and virtual resources required to instantiate a scenario. Virtual Machines (VMs) and containers can be used to reproduce hosts and services, while orchestration mechanisms support automated provisioning, configuration, and management of the training environment. Depending on the CTR architecture, these resources may be hosted locally, in the cloud, or across distributed infrastructure (Kampourakis et al. 2025; Lillemets et al. 2025).

Target Environment: The target environment represents the systems and technologies on which the exercise is conducted. Depending on the training objective, this may include Information Technology (IT), Operational Technology (OT), Industrial Control Systems, Internet of Things devices, or combinations of these environments. Simulation and emulation techniques can be used to reproduce the required systems and operational behaviour within a controlled setting (Lillemets et al. 2025).

Teams and Participant Roles: CTR exercises can assign different roles according to the objectives and structure of the exercise. Common roles include red teams conducting offensive activities and blue teams responsible for defence. White-team members may control, supervise, and evaluate the exercise, while additional technical or support roles can be used to maintain the underlying training environment. The exact team structure varies across CTR implementations and exercise designs (Lillemets et al. 2025).

Monitoring and Data Collection: Monitoring mechanisms collect information about network traffic, system events, scenario progression, and participant actions during an exercise. These data support situational awareness, technical analysis, scoring, after-action review, and evaluation of trainee performance. They can also provide feedback for improving subsequent scenarios and form an important foundation for adaptive training approaches (Lillemets et al. 2025).

3.2 History of CTR

The evolution of CTRs has closely followed the increasing sophistication of cyber threats and the rapid advancement of computing technologies. Initially developed as controlled network security testbeds for military experimentation and Computer Network Operations (CNO), CTRs have progressively evolved into comprehensive platforms supporting cybersecurity education, research, workforce development, and operational readiness. Advances in virtualization, cloud computing, software-defined networking, containerization, orchestration, and Artificial Intelligence (AI) have significantly expanded their capabilities, enabling highly realistic, scalable, and adaptive cyber training environments. This evolution demonstrates a clear transition from isolated experimental testbeds to intelligent cyber ecosystems capable of supporting complex multi-domain training scenarios (Davis and Magrath 2013; Priyadarshini 2018; Yamin et al. 2020; Kampourakis et al. 2025).

Early 2000s – Pre-range testbeds. The earliest CTRs evolved from network security testbeds and cyber experimentation laboratories that provided controlled environments for CNO, vulnerability assessment, attack simulation, and security experimentation. These early platforms established the foundation for realistic cybersecurity training by enabling researchers to safely evaluate attacks and defensive techniques without affecting operational systems (Davis and Magrath 2013).

2008 - 2013 - Emergence of national CTRs. During this period, governments, defence organisations, and research institutions began establishing dedicated CTRs to strengthen national cybersecurity capabilities. These platforms supported realistic cyber defence exercises, cyber warfare simulations, incident response training, and collaborative red-team/blue-team activities. This period marked the transition from

isolated security testbeds to organised cyber training environments designed to enhance operational readiness and workforce development (Davis and Magrath 2013; Priyadarshini 2018).

2014 - 2018 - Expansion and broader adoption. CTRs experienced widespread adoption across academia, government, military, and industry as organisations increasingly recognised the need for practical cybersecurity education and workforce development. During this period, advances in virtualization, cloud computing, software-defined networking (SDN), and automated infrastructure provisioning enabled the development of scalable, cost-effective, and highly realistic cyber training environments. These technological advances significantly improved the accessibility, repeatability, and operational realism of cyber range exercises while supporting larger and more diverse training scenarios (Priyadarshini 2018; Ukwandu et al. 2020).

2019 - 2024 - Intelligent and cloud-native CTRs. During this period, CTRs evolved into intelligent, cloud-native platforms that integrated virtualization, containerization, cloud computing, orchestration, and software-defined networking to support scalable and distributed cybersecurity training. Research increasingly focused on automating cyber range deployment, improving interoperability, incorporating CTI, and enhancing scenario management to create more realistic and adaptive training environments. These developments enabled CTRs to support complex cyber defence exercises, collaborative training, and large-scale experimentation while improving operational flexibility and resource efficiency (Ukwandu et al. 2020; Zhang et al. 2023; Stamatopoulos et al. 2024).

2025–2026 – More recently, AI-supported approaches, including LLM-based generation, RAG-enhanced LLM applications, and emerging agentic systems, have begun to influence scenario generation, CTR management, and adaptive training, signalling a transition towards more intelligent and adaptive cyber training environments (Kampourakis et al. 2025; Lillemets et al. 2025; Steininger et al. 2025; Lupinacci et al. 2025; Palumickas et al. 2026).

The evolution of CTRs demonstrates a clear transition from isolated security testbeds to intelligent, scalable, and interoperable cyber training ecosystems. Advances in virtualization, cloud computing, SDN, containerization, orchestration, and AI have progressively enhanced the realism, flexibility, and automation of cyber range environments. As cyber threats continue to evolve in sophistication and scale, modern CTRs are increasingly expected to support adaptive scenario generation, continuous learning, collaborative training, and AI-assisted decision making. Consequently, CTRs have become indispensable platforms for cybersecurity education, workforce development, cyber resilience assessment, and security research across academia, industry, government, and defence sectors (Kampourakis et al. 2025; Lillemets et al. 2025; Steininger et al. 2025; Garg et al. 2025).

3.3 Significance of Cyber Training Ranges

CTRs have become indispensable platforms for strengthening cybersecurity preparedness by providing realistic, repeatable, and controlled environments for education, training, experimentation, and operational validation. Unlike traditional classroom-based learning or theoretical exercises, CTRs enable participants to acquire practical experience through immersive, scenario-based exercises that closely emulate real-world cyber-attacks, defensive operations, and incident response activities. This experiential learning approach enhances technical competencies, decision-making, teamwork, and situational awareness while allowing organisations to safely evaluate cyber defence strategies, security technologies, and operational procedures without exposing production environments to unnecessary risks (Priyadarshini 2018; Ukwandu et al. 2020; Zhang et al. 2023).

Modern CTRs extend beyond technical training by providing integrated environments that support the complete lifecycle of cybersecurity exercises, including scenario creation, deployment, execution, monitoring, performance assessment, and post-exercise analysis. Their structured architectures enable organisations to evaluate security technologies, validate incident response procedures, measure workforce readiness, and improve organisational cyber resilience in a controlled and repeatable manner. Recent research also highlights the growing need for cyber ranges to support cloud-native infrastructures, cyber–physical systems, intelligent automation, interoperability, and scalable scenario management in response to increasingly sophisticated cyber threats (Ukwandu et al. 2020; Zhang et al. 2023; Stamatopoulos et al. 2024; Kampourakis et al. 2025; Lillemets et al. 2025). Consequently, CTRs have evolved from standalone training platforms into strategic cybersecurity infrastructures that support education, research, capability development, and continuous organisational preparedness.

Although significant advances have been made in the design, architecture, and deployment of CTRs, the educational effectiveness of these platforms ultimately depends on the quality, realism, and adaptability of the training scenarios they provide. Regardless of the underlying infrastructure or technology stack, static or poorly designed scenarios limit learner engagement, reduce realism, and fail to adequately prepare trainees for rapidly evolving cyber threats. Consequently, recent research has increasingly focused on intelligent scenario generation, automation, CTI integration, and AI-supported approaches, including the use of LLMs to develop more adaptive, scalable, and realistic cyber training environments (Russo et al. 2018; Hannay et al. 2021; Zacharis et al. 2024; Garg et al. 2025). The following section reviews the evolution of cyber range scenario generation and critically examines the transition from traditional manual approaches to intelligent AI-driven techniques.

4 Scenario Generation in Cyber Ranges

Scenario generation involves the design and construction of CTR exercises by defining the training environment, participant roles, events, threats, rules, objectives, and injects that determine how an exercise will progress. Scenarios may be static or dynamic.

Static scenarios generally follow predefined events or sequences, whereas dynamic scenarios can change in response to trainee actions, system states, or newly introduced threat information. Scenario development can therefore range from manual and rule-based approaches to more automated methods incorporating CTI and AI-supported techniques, including LLM-based generation (Yamin et al. 2020; Garg et al. 2025; Yamin et al. 2024).

4.1 Significance of Scenario Generation in CTRs

Scenario generation is crucial because the quality of a training scenario directly impacts the realism, relevance, and educational value of a CTR exercise. Well-crafted scenarios enable trainees to practise responding to realistic attacks in a controlled setting, avoiding risks to actual systems. They also assist instructors in establishing clear training goals, assigning roles, managing exercise difficulty, and evaluating performance. These scenarios serve various purposes, such as developing technical skills, incident response, team coordination, decision-making, and organizational readiness. Consequently, effective scenario creation is vital to ensure CTR exercises accurately simulate real-world threats and deliver valuable learning experiences (Yamin et al. 2020; Leitner et al. 2020; Glas et al. 2023).

4.2 Scenario Generation Lifecycle

Scenario generation involves multiple interconnected stages that transform training requirements into an executable CTR exercise. Although CTR platforms may differ in how this process is implemented, the lifecycle generally includes defining training objectives, designing the technical environment, assigning participant roles, selecting threats and attack events, validating and deploying the scenario, monitoring its execution, and assessing trainee performance (Russo et al. 2018; Garg et al. 2025; Russo et al. 2020; Yamin and Katt 2022). The lifecycle begins by identifying the required skills and learning objectives. A suitable scenario is then designed by defining the network topology, systems, participant roles, attack events, and expected outcomes. Before deployment, the scenario should be validated for technical correctness and alignment with the intended training objectives (Russo et al. 2018). During execution, trainee actions and system events can be monitored and recorded for assessment. The resulting performance data can subsequently inform improvements to future scenarios and, in more advanced approaches, support adaptive scenario generation (Hannay et al. 2021; Gizelis et al. 2025). Recently, Pfaller et al. proposed a process-based cyber exercise lifecycle in which process models support scenario planning, process engines automate scenario execution and adaptation, and execution logs provide information for subsequent evaluation (Pfaller et al. 2025).

Fig. 4 presents the general scenario-generation lifecycle used to organise the discussion in the following subsections.

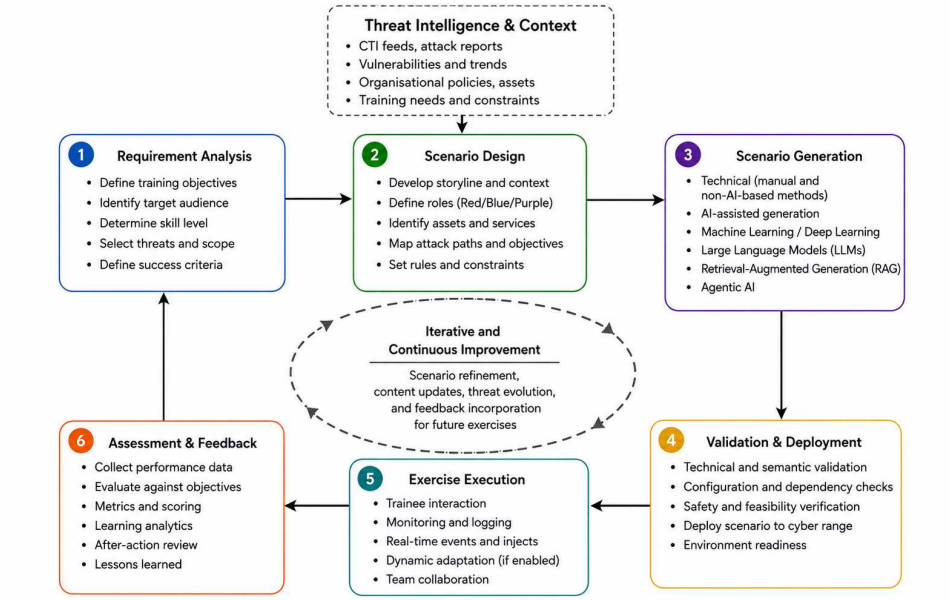


Fig. 4: General scenario generation lifecycle for cyber training ranges. The lifecycle begins with training requirements and threat analysis, followed by scenario design, generation, validation and deployment, exercise execution, and assessment. Lessons learned and performance feedback are incorporated into subsequent iterations, enabling continuous refinement of future scenarios.

4.3 Taxonomy of CTR Scenario Generation

Based on the systematic analysis of the reviewed literature, this study organises CTR scenario-generation approaches into two principal categories (Fig. 5): non-AI-based and AI-based approaches. Non-AI-based approaches include simulation-based, template-based, rule-based, attack-graph-based, and Model-Driven Engineering (MDE)-based methods, which primarily rely on predefined models, rules, specifications, or deterministic automation for scenario construction and deployment. AI-based approaches are further classified into traditional AI and LLM-based methods. Traditional AI encompasses ML, including DL, generative models such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), and RL, whereas LLM-based approaches include standalone LLMs, RAG-enhanced LLMs, and emerging agentic AI approaches (Braghin et al. 2019; Yamin et al. 2020, 2024; Garg et al. 2025; Lupinacci et al. 2025).

In addition to classifying the underlying generation technologies, the taxonomy identifies AI-generated and adaptive attack scenarios as a cross-cutting capability. In these approaches, AI may generate, select, or adapt attacker actions according to the state of the training environment, trainee or defender responses, and evolving scenario conditions (Agrawal et al. 2026; Palumickas et al. 2026). Consequently, AI-driven attack generation may be realised through different technologies, particularly RL, LLM-based

generation, and agentic approaches, rather than representing a separate AI technique (Yamin et al. 2024; Lupinacci et al. 2025; Palumickas et al. 2026).

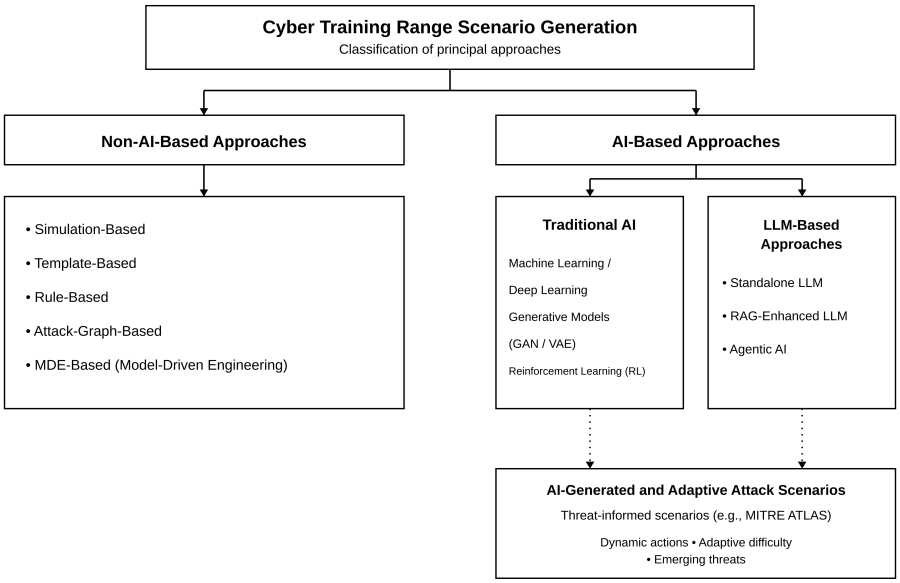


Fig. 5: Taxonomy of cyber training range scenario-generation approaches. Solid arrows represent hierarchical relationships, while dotted arrows indicate AI-enabled attack-generation capabilities shared across AI-based approaches.

4.4 Key Modules of Scenario Generation in Cyber Training Ranges

Scenario generation in CTRs can be implemented through several approaches that differ in how scenario content, infrastructure, attack behaviour, and exercise progression are represented and generated. Based on the reviewed literature, the principal approaches can be broadly grouped into simulation-based, template-based, rule-based, attack-graph-based, model-driven engineering, and AI-supported methods (Yamin et al. 2020; Garg et al. 2025). These approaches are not necessarily mutually exclusive; practical CTR implementations may combine elements from multiple approaches, such as structured scenario models, automated deployment, randomisation, and AI-supported content generation.

As illustrated by the general scenario-generation lifecycle presented earlier, a CTR scenario progresses from training requirements and design through generation, validation, deployment, execution, and assessment. While that lifecycle describes the stages through which a scenario progresses, it does not specify the internal information

that must be represented within the scenario itself. Simulation-based scenario generation is therefore examined first because it provides a useful conceptual structure for understanding these internal elements. Yamin et al. identify six principal modules in the scenario preparation process: story, entity, infrastructure, conditions, events, and Terminal Learning Objectives (TLOs) (Yamin et al. 2024).

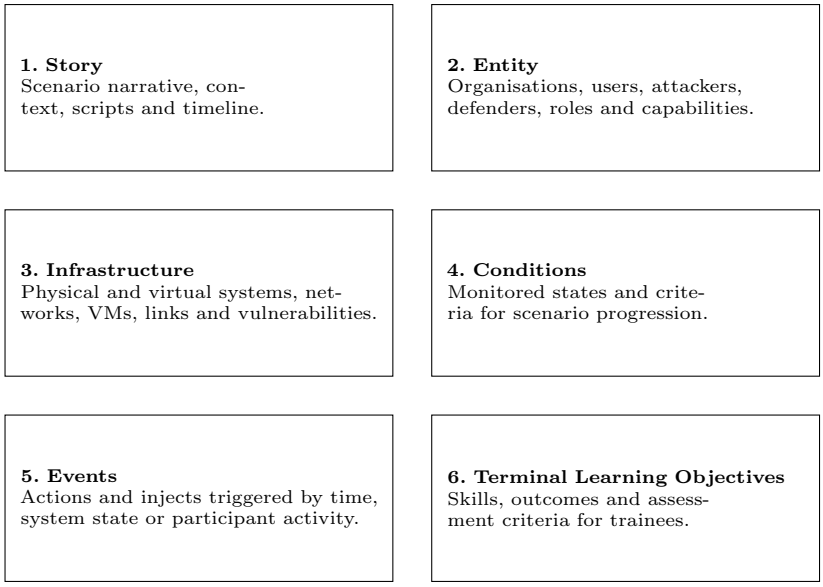


Fig. 6: Six modules of simulation-based scenario generation, adapted from Yamin et al. (Yamin et al. 2024).

Figure 6 summarises the six principal modules used to structure the preparation of a simulation-based cybersecurity exercise. Together, these modules connect the narrative and participating entities with the technical environment, runtime conditions, exercise events, and intended learning outcomes. Their functions are described below (Yamin et al. 2024).

4.4.1 Story

The story defines the overall context and narrative of the cybersecurity exercise. It establishes the situation being simulated, the organisations or systems involved, and the security incident around which the exercise is constructed. For example, a scenario may describe a ransomware campaign against a financial institution in which attackers initially compromise a third-party supplier, obtain access to the internal network, escalate privileges, and attempt to encrypt critical systems.

A story may contain several scripts or sub-stories that represent different parts or perspectives of the overall incident. These scripts organise a sequence of events and can specify when particular activities occur and how quickly the scenario progresses.

For example, one script may represent the initial phishing compromise, while another may represent lateral movement and a later data-exfiltration attempt. This structure allows a complex cyber incident to be represented as related stages while maintaining a coherent overall narrative (Yamin et al. 2024).

4.4.2 Entity

Entities represent the actors and organisational components that participate in or are affected by the scenario. They may include organisations, departments, users, attackers, defenders, system administrators, third-party providers, or other roles required by the exercise. Each entity is associated with responsibilities and capabilities that determine how it can interact with the scenario.

For example, in a ransomware exercise, a red-team entity may attempt to gain unauthorised access and move laterally through the network, while a blue-team entity is responsible for detecting suspicious activity, isolating affected systems, and initiating incident-response procedures. An organisational entity may also define policies, available tools, authority levels, and relevant skills that constrain how participants respond to the simulated incident (Yamin et al. 2024).

4.4.3 Infrastructure

Infrastructure defines the physical and virtual resources required to execute the scenario. These resources may include VMs, servers, workstations, switches, networks, applications, services, and vulnerabilities associated with the participating entities. Infrastructure therefore translates the scenario narrative into the technical environment on which the exercise can be executed (Yamin et al. 2024).

For example, a ransomware scenario may contain a Windows domain controller, an employee workstation, a vulnerable web server, a file server, a firewall, and an attacker VM connected through several network segments. The model described by Yamin et al. includes three infrastructure parameters: *count*, which specifies the number of VMs and switches; *dependencies*, which identify nodes that depend on other nodes and therefore influence deployment order; and *links*, which describe switch or node relationships (Yamin et al. 2024).

4.4.4 Conditions

Conditions represent measurable states of the training environment that can be monitored during scenario execution. They are used to determine whether particular criteria have been satisfied and can therefore influence scoring or trigger subsequent exercise events. In the scenario model described by Yamin et al., conditions are associated with VMs and can return a value between 0 and 1 representing the monitored state (Yamin et al. 2024).

For example, a condition could periodically check whether a malicious process is running on a compromised workstation. The condition may specify a *command* that executes a monitoring script, an *interval* defining how frequently the command is executed, and a *source* identifying the package containing the required monitoring artefact. A value of 1 could indicate that the malicious process is present, causing

the associated scenario event to become eligible for activation. Similarly, a condition could monitor whether a defender has successfully isolated a compromised host before allowing the scenario to progress to the next stage.

4.4.5 Events

Events define actions or developments that occur during the exercise. They are associated with rules or conditions that determine when particular activities or injects should be activated. Events may represent attacker actions, organisational responses, technical changes, or information supplied to participants. They therefore provide the mechanism through which the scenario progresses during execution (Yamin et al. 2024).

For example, a rule may specify that if a phishing attachment is executed on the employee workstation, the scenario should trigger an inject representing command-and-control communication. A subsequent rule may state that if the blue team fails to isolate the affected host within ten minutes, a lateral movement event is activated. Conversely, successful isolation could trigger an alternative branch in which the attacker attempts another entry path.

Injects provide the concrete actions associated with events. They may execute a technical action between entities or provide contextual information, such as an email, security alert, help-desk message, or simulated news report. Yamin et al. describe injects as being associated with a source, relevant TLOs, capabilities, descriptive information, and environment parameters needed to execute the corresponding package (Yamin et al. 2024).

4.4.6 Terminal Learning Objectives

TLOs define the specific knowledge, skills, and capabilities that participants are expected to demonstrate by completing the exercise. Linking scenario events and injects to TLOs helps ensure that technical scenario activity is connected to an explicit educational purpose rather than being included only for realism or complexity (Yamin et al. 2024).

For example, a defensive ransomware scenario may define TLOs requiring a trainee to detect malicious activity from security logs, identify the compromised endpoint, isolate the affected system, preserve relevant forensic evidence, and communicate the incident according to organisational response procedures. Each TLO can include a name and description, while an evaluation component defines the criteria and metrics used to determine whether the objective has been achieved. The model described by Yamin et al. also supports a minimum score for determining successful completion (Yamin et al. 2024).

Although this module-based structure provides a clear and systematic way to represent scenario content, it does not by itself solve the problem of generating realistic and adaptive exercises. Storylines, entities, infrastructure, conditions, events, and TLOs must still be defined, configured, and validated, which can require considerable subject-matter expertise and preparation effort. The quality of the resulting exercise also depends on whether the technical infrastructure can actually support the events and vulnerabilities specified in the scenario.

A further limitation is that predefined scripts, conditions, and events can constrain scenario behaviour to paths anticipated during design. Although conditions and injects support dynamic progression, the model does not automatically guarantee adaptation to previously unseen trainee actions or newly emerging threats. In addition, dependencies between narrative, infrastructure, events, and learning objectives create a validation problem: a scenario may be logically coherent while containing technical actions that cannot be executed on the target CTR, or it may be technically valid without adequately measuring the intended TLOs. These limitations motivate the use of more automated approaches, including templates, rules, attack graphs, model-driven engineering, and AI-supported generation, which are examined in the following subsection (Yamin et al. 2024; Garg et al. 2025).

4.5 Approaches for Generating Cyberattacks/Cyber Range Scenarios

Beyond the simulation-oriented structure described in Section 4.4, the reviewed literature demonstrates several technical approaches for constructing and automating CTR scenarios. These approaches can be broadly discussed as template-based, rule-based, attack-graph-based, model-driven engineering, and AI-based generation. The categories are not mutually exclusive. For example, a model-driven scenario may use rules for validation, an AI-supported system may use an attack graph for attack planning, and a template-based system may incorporate randomisation during scenario instantiation. Consequently, these approaches are distinguished primarily by the mechanism used to represent, construct, or adapt scenario content rather than as completely independent classes (Garg et al. 2025). Ribaudo and Valenza provide an example of a semi-automatic approach based on Flow-Based Programming and Node-RED, where reusable modules can be combined to generate cybersecurity exercises with different vulnerabilities (Ribaudo and Valenza 2019).

4.5.1 Template-Based Generation

Template-based generation constructs scenarios from predefined and reusable scenario structures. A template can specify elements such as network topology, VM types, services, vulnerabilities, attack components, and configuration parameters. Rather than designing every exercise from the beginning, an instructor selects or modifies an existing template and supplies the parameters required to instantiate a particular training environment. Parameterisation allows the same underlying structure to be reused with different hosts, network configurations, vulnerabilities, or exercise settings.

Template-based generation improves repeatability and reduces scenario preparation effort because validated components can be reused across multiple exercises. It can also support automated deployment when templates are translated into infrastructure configurations. However, its flexibility is limited by the scenarios and components anticipated when the templates were created. Substantial changes in threat behaviour, infrastructure, or learning objectives may therefore require templates to be manually modified or new templates to be developed. Excessive reuse can also make exercises predictable unless parameterisation or randomisation is incorporated (Garg et al. 2025).

4.5.2 Rule-Based Generation

Rule-based generation represents scenario behaviour using predefined logical rules, conditions, and actions. A rule typically specifies that when a particular condition or system state is observed, a corresponding scenario action should occur. For example, a rule may state that if a trainee fails to isolate a compromised workstation within a specified period, the scenario activates a lateral-movement event. Alternatively, successful containment may cause the exercise to proceed through a different branch.

This approach provides deterministic control over scenario progression and allows instructors to encode known attack-response relationships and pedagogical requirements. Rules can also be combined with monitoring data to create branching or partially dynamic exercises. However, rule-based systems depend heavily on expert-authored rules and generally respond only to situations anticipated during scenario design. As the number of conditions, events, and possible interactions increases, the rule base can become difficult to construct, maintain, and validate. This limits its ability to respond autonomously to previously unseen trainee behaviour or emerging threats (Garg et al. 2025; Yamin et al. 2024). A practical example is CyPROM, proposed by Beuran et al., which automates scenario progression through predefined triggers, actions, and branching. The next scenario step is selected based on the outcome of the current action, allowing the exercise to follow different predefined paths in response to trainee activity (Beuran et al. 2019).

4.5.3 Attack-Graph-Based Generation

Attack-graph-based approaches represent possible attack progression as a graph in which nodes describe security states, systems, vulnerabilities, or attacker objectives, while edges represent transitions enabled by exploits or other attack actions. Scenario-generation mechanisms can use these graphs to identify feasible attack paths from an initial state towards a target objective. For example, an attack path may represent initial compromise of a public-facing server, privilege escalation, lateral movement to an internal host, and eventual access to a protected database.

Attack graphs provide a structured representation of dependencies between vulnerabilities and attack steps and can therefore support the generation or validation of technically feasible attack sequences. They are particularly useful when scenario objectives depend on specific attack paths. However, their construction depends on sufficiently accurate information about the target environment, vulnerabilities, connectivity, and attack prerequisites. Large infrastructures can also produce complex graphs containing many possible paths. Furthermore, an attack graph primarily represents technical attack progression and does not by itself capture all pedagogical, organisational, narrative, or trainee-assessment requirements of a complete CTR exercise (Garg et al. 2025).

4.5.4 Model-Driven Engineering-Based Generation

Model-driven engineering (MDE) represents a scenario at a higher level of abstraction and then transforms that representation into deployable infrastructure and executable scenario artefacts. Domain-Specific Languages (DSLs) can be used to describe scenario

elements such as systems, networks, services, vulnerabilities, and relationships in a structured and machine-readable form. Braghin et al., for example, propose the Cybersecurity Training Language (CTL), developed using Xtext, to describe a cyber range scenario. The resulting model is validated and automatically transformed through an intermediate representation into a HEAT template, which can then instantiate the required environment on OpenStack (Braghin et al. 2019). Similarly, Costa et al. proposed the Virtual Scenario Description Language (VSDL), a domain-specific language for specifying high-level cyber-range infrastructure requirements. VSDL translates scenario specifications into formal constraints that a Satisfiability Modulo Theories solver checks and, when valid, automatically converts the resulting model into deployment scripts for the virtual infrastructure (Costa et al. 2022). Somarakis et al. proposed a model-driven approach that uses a Security Assurance Model to represent an organisation, its assets, security posture, threats, trainee actions, and training objectives, enabling the generation of tailored cyber-range scenarios and supporting their automated deployment through simulation and emulation (Somarakis et al. 2019).

Other model-driven approaches extend this concept by combining scenario description with formal validation. Scenario models can be transformed into formal representations to verify properties such as attack-path feasibility or model consistency before deployment. This provides an important advantage over purely manual scenario construction because errors can potentially be identified before computing resources are instantiated. Specialised scenario description languages can also support modularity and reuse of scenario components.

The principal limitations are the effort required to design and maintain the DSL, metamodel, transformation rules, and platform-specific deployment mechanisms. Interoperability is also challenging because different CTRs may use different scenario representations and orchestration technologies. Specialised SDLs can provide richer semantics and automation than generic formats such as JavaScript Object Notation (JSON), YAML Ain't Markup Language (YAML), or Extensible Markup Language (XML), but their additional complexity can restrict adoption and cross-platform reuse (Braghin et al. 2019; Garg et al. 2025).

4.5.5 AI-Based Generation

AI-based generation shifts part of the scenario-construction process from explicitly predefined templates and rules towards data-driven or generative mechanisms. Depending on the technique, AI can support activities such as extracting scenario elements from CTI, generating attack narratives, constructing exercise content, selecting or adapting attack actions, and modifying scenarios according to contextual or trainee information. More recent approaches use LLMs to generate scenario components from natural-language requirements (Yamin et al. 2024; Garg et al. 2025).

The CyExec approach described by Yamin et al. illustrates this transition. Its scenario-design process uses two LLMs with different perspectives, including organisational and cybersecurity expertise, and incorporates RAG to provide additional contextual information during generation. The generated scenario is progressively developed around elements such as the story, entities, infrastructure, events, conditions,

and TLOs (Yamin et al. 2024). AI-supported approaches therefore have the potential to reduce manual authoring effort and produce more varied and context-sensitive scenarios.

However, AI-generated scenarios introduce additional validation requirements. LLMs can hallucinate, produce technically inconsistent information, or generate content that does not align with the capabilities of the target CTR or the intended learning objectives. AI-based generation can also reduce instructor control when generated or autonomous behaviour becomes less predictable. Consequently, generated scenarios require technical, pedagogical, and safety validation before deployment (Yamin et al. 2024; Garg et al. 2025).

4.5.6 MITRE ATT&CK-Based Attack Scenario Generation

MITRE ATT&CK provides a structured knowledge base of adversary tactics and techniques that can be used to represent realistic attack behaviour. In cyber training ranges, ATT&CK techniques can be mapped to scenario actions and attack sequences, providing a common structure for developing threat-informed exercises. Cyber Threat Intelligence (CTI)-driven generation pipelines can also use ATT&CK information to translate known adversary behaviours into scenario content. This approach is particularly applicable to traditional ICT systems and networks, where ATT&CK-aligned attack sequences can be used to create realistic training scenarios (MITRE 2026).

4.5.7 MITRE ATLAS-Based Attack Scenario Generation

MITRE ATLAS extends the threat modelling concepts of MITRE ATT&CK to address adversarial threats that specifically target Artificial Intelligence (AI) and Machine Learning (ML) systems. It provides a structured taxonomy of tactics and techniques associated with attacks against AI and ML systems, including attacks targeting different stages of the ML lifecycle. ATLAS also covers emerging threats involving Large Language Models (LLMs) and agentic AI systems. In cyber training ranges, ATLAS can therefore provide a useful basis for designing attack scenarios that target AI models, AI-enabled applications, and their supporting systems and infrastructure (Sachpelidis-Brozos et al. 2026).

4.5.8 Comparison and Need for Hybrid Approaches

No single scenario-generation approach addresses all requirements of a modern CTR. Template- and rule-based methods provide repeatability and instructor control but are constrained by predefined content. Attack graphs provide structured representations of feasible adversarial progression but focus primarily on technical attack paths. Model-driven engineering improves formalisation, validation, reuse, and automated deployment, although it introduces modelling complexity and interoperability challenges. AI-based methods offer greater flexibility, variation, and potential adaptation but introduce uncertainty and require stronger validation and human oversight.

The reviewed studies show that various scenario-generation methods target different phases of the CTR scenario lifecycle. Structured and model-driven approaches support

scenario specification, infrastructure modelling, and deployment (Somarakis et al. 2019; Costa et al. 2022), while CTI and attack-graph-based approaches provide structured information for representing threats and attack paths (Zacharis et al. 2024). AI-based methods extend these capabilities by supporting scenario generation and adaptation (Yamin et al. 2024). Based on these complementary capabilities, a hybrid strategy could combine instructor-defined learning objectives and operational constraints with structured models, threat information, and AI-assisted generation. Monitoring and assessment data could then support subsequent scenario refinement.

5 CTR Scenario Generation Approaches

As shown in Fig. 5, CTR scenario-generation approaches are broadly classified into two classes: (i) non-AI-based and (ii) AI-based approaches. These classes are presented in the following sections:

5.1 Non-AI-Based Scenario Generation Approaches

Before the recent adoption of AI, research on CTR scenario generation primarily focused on reducing the effort required to construct, configure, validate, and deploy cybersecurity exercises through structured representations and deterministic automation. These approaches include expert-authored scenarios, reusable templates, randomised scenario construction, DSLs, model-driven engineering, formal validation, and infrastructure orchestration. Although they do not learn or generate content using AI models, they established many of the mechanisms on which current automated and AI-supported CTRs depend. Representative approaches are discussed below in chronological and logical progression (Yamin et al. 2020; Garg et al. 2025).

5.1.1 CyRIS: Security-Requirement-Driven Scenario Generation

An early step towards automating scenario construction was CyRIS, which was designed to generate cybersecurity training environments from instructor- specified security requirements. Rather than manually configuring every host, the instructor describes the required training environment and security properties, and CyRIS automatically creates and configures the corresponding virtual environment. This approach reduces repetitive configuration effort while allowing the instructor to retain control over the intended exercise requirements (Pham et al. 2016).

The main contribution of CyRIS is therefore the translation of predefined security requirements into reproducible training environments. Its requirement-driven model improves repeatability and reduces the amount of manual configuration required for scenario instantiation. However, the scenario content and required security properties must still be specified in advance by the exercise designer. Consequently, automation primarily concerns environment generation and configuration rather than autonomous creation or adaptation of the complete training narrative. This limits the ability of the approach to respond dynamically to previously unseen trainee behaviour or rapidly changing threat information (Pham et al. 2016; Garg et al. 2025).

5.1.2 SecGen: Template-Based Randomised Generation

SecGen extended non-AI automation by introducing randomisation into scenario construction. It uses an XML-based configuration language and a modular library containing elements such as vulnerabilities, services, networks, users, and content. Scenario-specific logic selects and combines these components, while technologies such as Vagrant and Puppet provision the required VMs. This allows multiple variations of an exercise to be generated from reusable components rather than requiring instructors to manually build each environment from the beginning (Grimaldi et al. 2023).

The principal advantage of SecGen is variation. Randomising selected vulnerabilities, services, and configuration elements can reduce the predictability associated with repeatedly using an identical exercise and can increase scenario reuse. However, the generated variation remains constrained by the modules, dependencies, and logic already available in the scenario library. Randomisation therefore increases diversity but does not provide semantic reasoning about emerging threats, trainee performance, or the pedagogical suitability of newly generated combinations. Maintaining and extending the underlying module library also remains an expert-driven task (Grimaldi et al. 2023; Garg et al. 2025).

5.1.3 Model-Driven Scenario Deployment

Braghin et al. proposed a model-driven approach intended to reduce the gap between an abstract cybersecurity scenario and its deployable CTR infrastructure. Their framework introduces the CTL, a DSL developed using Xtext, through which an exercise designer can describe the simulated system at a higher level of abstraction. A validator checks the description against defined constraints, after which the model is transformed through an intermediate representation into a HEAT template that can be deployed on OpenStack (Braghin et al. 2019).

This approach improves reusability and consistency because infrastructure is defined through a structured model rather than repeatedly configured by hand. Automated transformation also reduces deployment effort and can detect some configuration problems before resources are instantiated. Nevertheless, the approach requires a specialised DSL, transformation rules, validation logic, and platform-specific deployment mechanisms. The abstraction does not remove the need for expert scenario design, and interoperability can become difficult when another CTR uses different models, orchestration technologies, or infrastructure semantics (Braghin et al. 2019).

5.1.4 CRACK: Model-Based Generation and Formal Validation

The CRACK approach further develops model-driven scenario engineering by combining scenario description, automated deployment, and formal validation. Instead of relying solely on manual trial-and-error configuration, an abstract scenario model can be transformed into both an executable DSL representation and a formal model. The formal representation can then be analysed before deployment to verify relevant scenario properties, while the executable representation supports infrastructure orchestration and cybersecurity operations (Russo et al. 2018).

An important advantage of this approach is that validation is incorporated into the scenario-engineering process. This can identify inconsistencies before an exercise is deployed and can improve reproducibility compared with manual configuration. The modular design also allows scenario components to be reused or modified. However, formal verification is limited to properties that can be represented by the underlying model and validation rules. Furthermore, model transformations and deployment artefacts must remain consistent with the target infrastructure. Changes to platforms, vulnerabilities, or scenario requirements may therefore require corresponding changes to the models and transformation mechanisms (Russo et al. 2018).

5.1.5 CRATE and Infrastructure-Oriented Automation

Subsequent CTR research increasingly incorporated orchestration to automate the provisioning and management of larger training environments. CRATE, for example, focuses on automating cyber-range infrastructure and supporting the deployment and management of exercise environments. Infrastructure automation reduces repetitive configuration and facilitates the recreation of complex networked environments, making exercises more reproducible and easier to reset between training sessions (Gustafsson and Almroth 2020).

This represents an important progression from individual scenario specification towards automation of the operational environment in which the scenario executes. However, infrastructure automation should not be equated with complete scenario generation. Orchestration can efficiently create, configure, and restore hosts and networks, but the exercise narrative, learning objectives, threat behaviour, event logic, and assessment criteria still require explicit specification. Consequently, orchestration solves an important deployment problem without independently solving the broader scenario-authoring problem (Gustafsson and Almroth 2020; Yamin et al. 2020).

5.1.6 Scenario Description and Interoperability

As CTRs became more automated, scenario description increasingly became an interoperability issue. Structured formats such as JSON, YAML, and XML can describe infrastructure and configuration information, while specialised SDLs provide richer representations of scenario-specific concepts. Formal or semi-formal scenario descriptions enable scenario elements to be processed by deployment and orchestration mechanisms and can improve reproducibility and reuse across exercises (Yamin et al. 2020; Garg et al. 2025).

However, no single scenario-description approach has become universally adopted across heterogeneous CTR platforms. A description that captures network topology may not adequately represent narrative progression, attack events, learning objectives, scoring, or runtime conditions. Conversely, highly specialised languages can represent richer scenario semantics but increase modelling complexity and may require platform-specific transformations. This creates an interoperability gap between scenario description, generation, and execution, particularly when scenarios are to be shared or migrated between different CTR platforms (Garg et al. 2025).

As summarised in Table 3, non-AI approaches progressively reduce manual configuration through requirement-driven generation, randomisation, structured modelling,

Table 3: Comparison of Representative Non-AI-Based Scenario Generation Approaches

Method / Approach and Coverage	Key Contributions	Key Limitation	Model Type	Future Work / Recommendation
CyRIS (Pham et al. 2016) Coverage: △	Generates reproducible training environments from predefined security requirements and reduces manual VM configuration.	Requires expert-defined requirements; limited autonomous scenario adaptation.	Requirement-driven automation	Extend generation beyond infrastructure configuration towards dynamic scenario and threat adaptation.
SecGen (Garg et al. 2025) Coverage: △	Uses reusable modules and randomisation to generate variations in vulnerabilities, services, users, networks, and VM configurations.	Variation is constrained by predefined modules and generation logic.	Template-based/randomised	Increase reusable component libraries and incorporate context-aware selection and validation.
CRACK (Russo et al. 2020) Coverage: ▲	Combines abstract scenario modelling, formal validation, executable scenario representation, and automated deployment.	Verification is constrained by the properties represented in the formal model and requires maintenance of model transformations.	Model-driven/formally validated	Extend semantic validation and improve reusable, platform-independent scenario representations.
Model-driven CTL (Somarakis et al. 2019) Coverage: △	Uses a DSL and model transformation to convert high-level scenario descriptions into deployable HEAT/OpenStack infrastructure.	Requires specialised modelling and platform-specific transformation mechanisms.	Model-driven engineering	Improve interoperability and portability across heterogeneous CTR platforms.
CRATE (Gustafsson and Almroth 2020) Coverage: △	Automates provisioning, configuration, and management of complex cyber-range infrastructure and supports repeatable exercises.	Primarily addresses infrastructure automation rather than autonomous generation of the complete exercise scenario.	Infrastructure orchestration	Integrate infrastructure automation more closely with scenario logic, learning objectives, and adaptive execution.
Structured SDL approaches (Arshad et al. 2021; Dalla Costa and Kuusijärvi 2022) Coverage: △	Provide machine-readable descriptions supporting scenario reuse, deployment automation, and integration between scenario components.	No universally adopted SDL; expressiveness and cross-platform interoperability remain inconsistent.	DSL/structured description	Develop interoperable scenario representations covering infrastructure, narrative, events, TLOs, and assessment.

Coverage: △ = partial coverage; ▲ = stronger coverage within the comparison. Note: Coverage ratings and recommendations reflect the authors' synthesis.

formal validation, and infrastructure orchestration. However, automation of deployment does not necessarily imply automation of scenario reasoning or adaptation. Several recurring limitations can therefore be identified across the reviewed approaches.

- **Preparation effort and expertise dependency:** Non-AI approaches continue to depend substantially on cybersecurity and CTR experts to define scenario requirements, templates, rules, models, infrastructure, and learning objectives. Manual configuration can be complex, time-consuming, and error-prone, while model-driven automation shifts part of this effort towards constructing and maintaining formal scenario descriptions and transformation mechanisms (Braghin et al. 2019; Russo et al. 2018; Yamin et al. 2020).
- **Limited adaptability:** Templates, predefined rules, scripts, and formal models generally operate within behaviour anticipated during scenario design. Randomisation can produce different configurations, but it does not independently reason about new threats or adapt exercise content to previously unseen trainee behaviour. Consequently, substantial changes in threats, objectives, or participant capability can require manual modification of the scenario artefacts (Grimaldi et al. 2023; Garg et al. 2025).
- **Validation and technical consistency:** Automated deployment reduces some configuration errors, and formal approaches such as CRACK provide mechanisms for validating scenario properties before execution. Nevertheless, validation is limited by the information and constraints represented in the model. A scenario may be structurally valid while still failing to capture all operational, pedagogical, or runtime requirements (Russo et al. 2018).
- **Predictability and constrained variation:** Repeated use of fixed scenarios can allow trainees to become familiar with attack sequences and expected solutions. Template-based randomisation, such as that used by SecGen, reduces this problem by varying selected scenario elements; however, the possible variations remain bounded by the predefined component library and generation logic (Grimaldi et al. 2023).
- **Interoperability and modelling fragmentation:** CTRs use different DSLs, data formats, orchestration technologies, and infrastructure platforms. As a result, scenarios developed for one environment may not be directly portable to another. Richer specialised models improve expressiveness but can also increase implementation complexity and platform dependency (Braghin et al. 2019; Garg et al. 2025).

Overall, non-AI approaches established essential foundations for automated CTR scenario generation by improving repeatability, randomisation, formal representation, validation, and deployment. Their principal limitation is not the absence of automation, since several approaches provide substantial automation, but the dependence on predefined human-authored knowledge, components, rules, and models. This distinction is important: non-AI automation can efficiently instantiate and vary what has already been specified, whereas it has limited capability to interpret new threat information, generate previously unspecified scenario content, or adapt scenario

behaviour autonomously. These limitations provide the motivation for the AI-based scenario-generation approaches examined in the following section.

5.2 AI-Based Scenario Generation

The transition towards AI-based scenario generation extends CTR automation beyond predefined templates, rules, and deterministic models. AI techniques can analyse threat information, identify relevant attack patterns, forecast emerging threats, generate synthetic threat data, control adaptive adversaries, and produce scenario content according to training requirements. For clarity, the approaches identified in the reviewed literature are divided here into two broad groups: (i) traditional AI (non-LLM)-based approaches, including ML, DL, generative models, RL, and multi-agent systems; and (ii) LLM-based approaches, including RAG and agentic architectures. This distinction is useful because the two groups differ substantially in their inputs, generation mechanisms, adaptability, and validation requirements (Garg et al. 2025; Zacharis et al. 2024; Yamin et al. 2024).

5.2.1 Traditional AI-Based Scenario Generation

In this review, traditional AI-based scenario generation refers to AI techniques that do not principally rely on LLMs. These include ML and DL models for threat analysis and forecasting, generative models such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), and RL-based agents for adaptive attack and defence behaviour. These approaches progressively extend non-AI automation by allowing scenario content or behaviour to be derived from data rather than being completely specified in advance.

AI-Driven Threat Forecasting: Zacharis et al. (Zacharis et al. 2024) integrated AI-driven sector-specific CTI and threat forecasting into the cybersecurity exercise-content generation lifecycle. Their approach analyses threat information and evaluates forecasting techniques, including Long Short-Term Memory (LSTM) networks and other time-series models, to identify current and emerging sector-relevant threats. The resulting information is incorporated into exercise-generation processes so that scenarios can better reflect threats that are likely to affect the target organisation or sector. The principal advantage is increased relevance: scenario selection is informed by observed and forecast threat trends rather than solely by instructor judgement or static threat libraries.

However, forecasting quality depends strongly on the quality, representativeness, and temporal coverage of the available data, including unexpected and future circumstances. LSTM models in particular may require substantial training data, computational resources, and hyperparameter tuning, while abrupt changes in attacker behaviour may remain difficult to forecast reliably. Threat forecasting can therefore determine *what* threats may be relevant, but it does not independently solve the complete problem of generating technically executable and pedagogically appropriate CTR scenarios.

GAN- and VAE-Based Threat Generation: A different form of AI-based scenario support uses generative ML models. GANs and VAEs are generative AI techniques, although they are fundamentally different from LLMs. Vadisetty and

Polamarasetti examined their use for creating synthetic cybersecurity threats, including malware variants, phishing content, and network-intrusion traffic. By learning statistical patterns from existing attack data, these models can produce additional samples that resemble known attacks while introducing variations that can be used to stress-test defensive mechanisms (Vadisetty and Polamarasetti 2024).

The major advantage of these approaches is the ability to generate large and diverse synthetic threat datasets, including variants that may be absent from a limited real-world training dataset. This can improve the variety of attack material available for scenario construction. However, synthetic realism does not automatically imply scenario realism. Generated malware, traffic, or phishing samples still need to be placed within a coherent CTR infrastructure, attack sequence, learning objective, and assessment process. The quality of the generated output is also dependent on the training data (the variation in the training data accurately reflects the expected generated data values) and the model's hyperparameter values, and generated artefacts require validation before operational use.

RL and Multi-Agent Adaptive Simulation: RL extends scenario generation from generating static content towards generating adaptive behaviour. Agrawal et al. (Agrawal et al. 2026) proposed an AI-driven Multi-Agent System (MAS) for adaptive cyberattack simulation and automated incident response. Their framework integrates the CICIDS2017 and UNSW-NB15 datasets with the CyDER 2.0 environment and uses RL and anomaly-detection techniques to enable attack and defence agents to modify their behaviour during scenario execution. This allows attacker–defender interactions to evolve dynamically rather than following an entirely predetermined attack script.

The main advantage of this approach is behavioural adaptation: an RL-enabled agent can modify actions according to changes in the environment and the behaviour of opposing agents. This provides a closer approximation of the interactive nature of real cyber conflict than fixed scripts. Nevertheless, the behaviour learned by an RL agent is influenced by its training environment, reward function, state representation, and available actions. An agent may therefore learn behaviour that achieves a reward without necessarily supporting the intended learning objective. Complex multi-agent behaviour can also be harder to reproduce and validate than deterministic exercise logic, requiring strong constraints and evaluation mechanisms (Garg et al. 2025).

The representative traditional AI-based approaches identified in this review are summarised in Table 4.

As shown in Table 4, traditional AI-based approaches introduce important capabilities that are largely absent from deterministic scenario generation. Forecasting improves threat relevance, generative models increase the diversity of available cyber-threat artefacts, and RL enables adaptive runtime behaviour. Nevertheless, these techniques generally address particular stages or components of the scenario lifecycle rather than generating and managing the complete scenario from natural-language requirements. This limitation provides a natural transition to LLM-based approaches, which can generate multiple interconnected scenario components through a common language-based interface.

Table 4: Comparison of Representative Traditional AI-Based Scenario Generation Approaches

Method / Approach and Coverage	Key Contributions	Key Limitation	ML Type	Future Work / Recommendation
Zacharis et al. (Zacharis et al. 2024) Coverage: △	Integrates sector-specific CTI and forecasting into exercise-content generation to improve scenario relevance to emerging threats.	Accuracy depends on data quality and forecasting performance; does not independently generate the complete executable scenario.	ML/LSTM/forecasting	Combine forecasting with automated scenario construction, infrastructure validation, and continuous CTI updates.
Vadisetty and Polamarasetti (Vadisetty and Polamarasetti 2024) Coverage: △	Uses generative models to synthesise malware, phishing content, and network-intrusion data for diverse threat simulation.	Synthetic threat realism does not guarantee complete scenario, pedagogical, or infrastructure validity.	GAN/VAE	Integrate synthetic threat generation with validated CTR scenario models, learning objectives, and deployment mechanisms.
Agrawal et al. (Agrawal et al. 2026) Coverage: ▲	Uses RL, anomaly detection, and MAS architecture for adaptive attacker and defender behaviour in a CTR.	Agent behaviour depends on training data, state/action design, and reward functions and can be difficult to reproduce and validate.	RL/MAS/anomaly detection	Strengthen explainability, reward alignment, safety constraints, and cross-range validation of adaptive agents.

Coverage: △ = partial coverage; ▲ = comparatively strong coverage of adaptive scenario-generation functions identified in this review.
Note: Coverage ratings and recommendations reflect the authors' synthesis.

5.2.2 LLM-Based Scenario Generation

LLM-based approaches represent a further change in scenario generation because they can interpret high-level natural-language requirements and generate multiple forms of scenario content from the same model. Unlike conventional ML models that are typically trained for a specific prediction or classification task, LLMs can generate narratives, attack events, infrastructure descriptions, participant tasks, TLOs, injects, and assessment material. Recent work has progressively extended this capability from LLM-assisted authoring towards RAG-grounded generation and agentic systems capable of performing retrieval, reasoning, validation, tool use, and deployment.

RAG is treated in this review as an augmentation mechanism for LLM-based generation rather than as a separate AI model or independent scenario-generation family, because it supplies external knowledge to the LLM during generation.

AI-Augmented Scenario Design: Reuter et al. examined the use of ChatGPT during the 2024 Austrian National Cybersecurity Exercise, where the model supported scenario design, infrastructure development, and content creation. The study shows that an LLM can assist exercise designers with early-stage ideation, drafting, and technical prototyping, reducing preparation effort and providing greater flexibility when scenario requirements change. However, the authors also identify important limitations, including hallucinations, limited temporal state awareness, and restricted usefulness for offensive simulation. Human expertise therefore remains necessary to validate the generated content and ensure that it is technically and operationally suitable for the exercise (Reuter et al. 2025).

CyExec and LLM-Based Exercise Generation: Yamin et al. (Yamin et al. 2024) introduced CyExec, an LLM-based approach for generating cybersecurity exercise scenarios. Their framework employs LLMs with different roles and uses RAG to provide additional contextual information during generation. The scenario is developed around elements such as the story, entities, infrastructure, events, conditions, and TLOs. The approach demonstrates that an instructor can move from high-level requirements towards a detailed exercise design without manually authoring every scenario component.

An important contribution of this work is its treatment of LLM *hallucination*. Rather than considering all novelty produced by an LLM to be undesirable, the authors investigate whether controlled generative variation can contribute to more creative cybersecurity scenarios. Nevertheless, hallucination remains a major technical risk because fabricated or inconsistent information may also generate unrealistic vulnerabilities, incorrect attack steps, or incompatible infrastructure. Novelty must therefore be distinguished from technical correctness, and expert validation remains necessary before generated scenarios are deployed.

Agentic AI extends LLM-based scenario generation from content creation towards multi-step task execution by combining capabilities such as planning, retrieval, tool use, intermediate evaluation, and iterative refinement.

Agent-Based Multi-Step Vulnerability Scenario Generation: Pavelic et al. investigated the use of LLMs in a multi-agent setup to generate Docker-based cyber range environments containing vulnerabilities that can be chained across networked systems. The approach explores the generation of multi-step scenarios involving activities such

as privilege escalation and lateral movement using GPT-4.1 and Mistral-7B. Their evaluation showed that GPT-4.1 produced more structured and coherent outputs, whereas Mistral-7B provided greater diversity in vulnerability chains but introduced more technical errors. Although the approach demonstrates the potential of LLM agents to automate the creation of complex multi-stage cyber range scenarios, the generated vulnerabilities were not validated for actual exploitability. This limits the technical assurance required before such scenarios can be deployed in operational training environments (Pavelić et al. 2025).

Agentic Vulnerability-Based Scenario Generation: Érsok et al. proposed a three-agent LLM framework for automatically generating cyber range scenarios from publicly available vulnerability information. The framework consists of a vulnerability analyzer that enriches Common Vulnerabilities and Exposures (CVE) information, a scenario architect that maps vulnerabilities to MITRE ATT&CK techniques and constructs multi-step attack flows, and a deployment generator that produces Ansible playbooks and Docker/Podman Compose files. The approach was evaluated using 28 high-severity CVEs across Mistral-Small 3.1, Qwen3.5, Llama 3.3, and Claude Sonnet. Compared with rule-based templates, the LLM-based approach increased ATT&CK technique diversity and produced more realistic multi-host topologies while generating structurally valid deployment artifacts. However, the authors identify limitations in semantic validation and deployment verification, indicating that further testing is required before fully autonomous scenario generation can be achieved (Érsok et al. 2026).

These studies show a progression from LLM-assisted scenario authoring towards multi-agent systems capable of specialised scenario-generation tasks. A further development is the integration of retrieval and agentic reasoning, enabling LLMs to draw on external knowledge and coordinate multiple steps during scenario generation and deployment.

ARCeR and Agentic RAG: An extension of LLM-based generation towards autonomous CTR definition and deployment through ARCeR was introduced by Lupinacci et al. (Lupinacci et al. 2025). ARCeR is an Agentic RAG system that accepts a high-level natural-language description of a desired CTR, retrieves relevant framework documentation, generates platform-specific configuration information, and supports subsequent range deployment. Unlike a basic RAG pipeline that primarily retrieves information and provides it as context to an LLM, the agentic architecture can perform multiple retrieval steps, plan intermediate operations, use external tools, and coordinate the generation process.

The principal advantage is that instructors do not need to manually master the syntax of each CTR configuration format. ARCeR can adapt to different range frameworks largely by changing the supporting documentation supplied to the retrieval component rather than retraining the underlying model. The authors report better performance than a pure RAG approach and successful generation for 90% of the tested scenarios.

However, ARCeR does not yet determine reliably whether a user's requested scenario exceeds the capabilities of the selected CTR platform. An unsupported topology or feature may therefore lead to a valid-looking configuration that fails during instantiation. The authors identify pre-generation capability validation and improved semantic correction as important future extensions (Lupinacci et al. 2025).

Thus, agentic reasoning improves automation but does not remove the requirement for technical validation.

ASCERT and Interactive RAG-Grounded Scenario Generation: Palumickas et al. (Palumickas et al. 2026) evaluated ASCERT as an interactive generative-AI approach for CTR scenario generation. ASCERT uses Llama-2-13B together with a ThreatIntelligenceRAG component, ChromaDB-based retrieval, MITRE ATT&CK references, a VirtualCISO component, and a crisis simulator. Rather than producing only a static narrative, the system generates successive adversary actions, presents role-aware response options to the trainee, evaluates the selected action, and creates branching scenario paths.

This system provides two important advantages. First, scenarios can vary between runs, reducing the predictability associated with fixed exercise scripts. Second, trainee decisions directly influence scenario progression, creating a more interactive training process. However, the evaluation also identifies significant limitations. Identical inputs can produce highly different outputs, limiting reproducibility for controlled evaluation. Some attack stages are skipped or incorrectly ordered, some generated actions are overly general, and several assessment metrics require further refinement.

The evaluation also demonstrates the importance of knowledge-source quality. The model reproduced real incident structure more accurately when the supporting dataset was smaller, focused, and carefully curated than when larger but less focused collections were supplied. This demonstrates that RAG does not benefit automatically from adding more documents; relevance and quality of the retrieved knowledge are critical to scenario fidelity.

The progression of representative LLM-based approaches is summarised in Table 5.

Tables 4 and 5 illustrate the progression from specialised, data-driven AI techniques to more integrated LLM-based approaches to scenario generation. Traditional AI techniques focus on specific capabilities, such as threat forecasting, synthetic threat generation, and adaptive attacker behaviour. By contrast, LLM-based approaches can bring multiple scenario-generation activities together within a common workflow. Retrieval-Augmented Generation (RAG) can improve the grounding of generated content in external knowledge sources, while agentic approaches can extend scenario generation towards automated configuration and deployment. However, increased automation and autonomy also introduce additional challenges involving knowledge quality, technical validity, reproducibility, pedagogical alignment, and human oversight.

6 Lessons Learned from AI-Based Scenario Generation

The comparative analysis in Tables 4 and 5 reveals that AI provides substantial improvements in scenario diversity, threat relevance, adaptability, and automation, but also introduces limitations that are different from those encountered in non-AI scenario generation. To systematically synthesise these observations, Table 6 summarises the principal research gaps with their open challenges. The following lessons are therefore derived from the strengths and limitations reported in all reviewed approaches.

1) Knowledge Quality Is More Important Than Knowledge Volume.

Table 5: Comparison of Representative LLM-Based Scenario Generation Approaches

Method / Approach and Coverage	Key Contributions	Key Limitation	LLM Type	Future Work / Recommendation
CyExec (Yamin et al. 2024) Coverage: △	Uses LLMs and RAG to generate structured cybersecurity exercise content, including narrative and scenario-design components.	Hallucination and technical inconsistency require expert validation before deployment.	LLM + RAG	Improve grounding, automated technical validation, and alignment between generated content and CTR capabilities.
ARCeR (Lupinacci et al. 2025) Coverage: ▲	Uses Agentic RAG to convert natural-language requirements into platform-specific CTR configuration and support automated deployment.	Cannot reliably determine beforehand whether requested features are supported by the target CTR platform.	LLM + Agentic RAG	Add capability-aware prevalidation, semantic correction, and broader cross-platform evaluation.
ASCERT (Palumickas et al. 2026) Coverage: ▲	Uses Llama-2-13B, RAG, threat intelligence, and interactive branching to generate decision-based scenarios that respond to trainee choices.	Limited reproducibility; attack stages can be skipped or misordered and accuracy depends heavily on curated source data.	Llama-2-13B + RAG	Improve stage enforcement, source curation, scoring metrics, reproducibility controls, and scenario validation.

Coverage: △ = partial coverage; ▲ = comparatively strong coverage of scenario-generation and management functions identified in this review. *Note:* Coverage ratings and recommendations reflect the authors' synthesis.

The effectiveness of AI-generated scenarios depends strongly on the quality and relevance of the information supplied to the generation process. This is particularly evident in the ASCERT evaluation, where focused and carefully curated knowledge sources produced scenarios that were more faithful to the corresponding real-world incidents than larger and less focused datasets. The finding indicates that simply increasing the quantity of CTI or RAG documents does not guarantee better scenario generation (Research Gaps 3 and 12 in Table 6). Retrieval, filtering, source quality, temporal relevance, and contextual suitability are therefore critical design considerations (Palumickas et al. 2026).

This lesson also applies to forecasting and conventional ML approaches. Models trained on incomplete, biased, outdated, or unrepresentative threat data can generate inaccurate predictions that subsequently influence scenario selection. Consequently, data curation should be treated as part of the scenario-generation lifecycle rather than as a one-time preprocessing task (Zacharis et al. 2024).

2) Generative Diversity Is Valuable but Must Be Controlled.

LLMs and other generative models can produce substantially more variation than manually scripted scenarios. This capability is valuable because trainees cannot simply memorise one predetermined attack sequence. LLMs can draw on patterns learned across broad cybersecurity information and, when combined with RAG, can incorporate additional organisation-specific or threat-specific knowledge to generate different scenario contexts, attack events, and participant challenges (Yamin et al. 2024; Palumickas et al. 2026). (Research Gaps 4, 11, and 13 in Table 6).

Yamin et al. additionally investigate whether controlled LLM hallucination can be used as a source of creative scenario variation (Yamin et al. 2024). However, creativity and correctness must not be conflated. A novel attack path may be educationally useful, whereas a fabricated vulnerability, impossible command, or technically inconsistent network configuration may invalidate an exercise. Generative diversity must therefore operate within technical and pedagogical constraints.

3) Adaptive Adversaries Can Improve Scenario Dynamism.

The RL and MAS approaches reviewed in Table 4 demonstrate that AI can support runtime adaptation rather than only pre-exercise content generation. Adaptive agents can modify attack or defence behaviour according to observed changes in the training environment, producing interactions that are less predictable than fixed scripts (Agrawal et al. 2026). (Research Gaps 5 and 7 in Table 6).

This capability can require trainees to analyse changing conditions and adjust defensive decisions rather than follow a memorised response sequence. However, greater agent autonomy also requires careful reward design, behavioural constraints, monitoring, and validation to ensure that agents continue to support the intended TLOs rather than merely optimise a numerical reward.

4) Grounding Improves LLM Generation but Does Not Guarantee Correctness.

RAG addresses an important limitation of standalone LLMs by supplying external cybersecurity knowledge during generation. This allows scenarios to incorporate information beyond the static knowledge encoded during model training. Nevertheless, the ASCERT results demonstrate that retrieved knowledge must be relevant and curated,

while ARCeR demonstrates that configuration generated from retrieved documentation can still be incompatible with the capabilities of the target CTR (Palumickas et al. 2026; Lupinacci et al. 2025). (Research Gaps 2, 3, and 12 in Table 6).

RAG should therefore be treated as a grounding mechanism rather than an automatic correctness mechanism. A robust CTR generation pipeline still requires source validation, semantic checks, infrastructure compatibility checks, and alignment with learning objectives.

5) Agentic AI Extends Generation Towards End-to-End Automation. The progression from standalone LLMs to RAG and then Agentic RAG is important for CTR automation. A conventional LLM primarily generates an output from a prompt, while RAG supplements the prompt with retrieved knowledge. Agentic systems can additionally plan intermediate tasks, perform repeated retrieval, invoke tools, evaluate intermediate results, and coordinate multiple generation steps. ARCeR illustrates this progression by extending natural-language scenario specification towards automated generation of CTR configuration files and deployment (Lupinacci et al. 2025). (Research Gaps 1, 2, and 9 in Table 6).

The results reported for ARCeR show that agentic orchestration can outperform a basic RAG pipeline for this task. However, its current inability to detect all unsupported user requests before generation illustrates that autonomous workflow execution must be accompanied by capability-aware validation. Increasing autonomy therefore increases, rather than eliminates, the need for verification mechanisms.

6) Reproducibility and Assessment Remain Open Challenges.

Generative AI introduces inherent variation into scenario generation. ASCERT produced different attack actions, user choices, and outcomes even when inputs were held constant. This dynamism is advantageous for reducing scenario predictability but complicates controlled experimentation, standardised assessment, and direct comparison of trainee performance (Palumickas et al. 2026). (Research Gaps 4, 6, and 11 in Table 6).

Future CTRs therefore need mechanisms that balance variation with controlled evaluation. Possible directions include fixing model parameters where appropriate, recording prompts and retrieved context, logging model and tool versions, constraining allowable scenario branches, and mapping generated events consistently to predefined TLOs and assessment metrics.

7) Human Oversight Remains Necessary.

Across the reviewed AI approaches, increased automation does not remove the need for cybersecurity and training expertise. ML forecasting requires interpretation of threat relevance; GAN- and VAE-generated artefacts require validation; RL agents require carefully designed environments and objectives; LLM-generated scenarios require checking for hallucinations and technical inconsistencies; and Agentic RAG systems require verification that generated configurations are supported by the target CTR (Zacharis et al. 2024; Vadisetty and Polamarasetti 2024; Agrawal et al. 2026; Yamin et al. 2024; Lupinacci et al. 2025). (Research Gaps 11, 12, and 13 in Table 6).

The evidence therefore supports a human-in-the-loop architecture rather than unrestricted autonomous scenario generation. AI is particularly valuable for reducing repetitive authoring effort, increasing scenario diversity, processing large bodies of threat

information, and adapting exercise behaviour, while instructors retain responsibility for validating technical feasibility, safety, pedagogical relevance, and assessment objectives.

8) Overall Lesson.

The central lesson from the reviewed literature is that the evolution from ML and DL towards generative models, LLMs, RAG, and Agentic AI progressively increases the scope of scenario-generation automation (Yamin et al. 2024; Reuter et al. 2025; Lupinacci et al. 2025; Pavelić et al. 2025; Érsok et al. 2026). However, no individual AI technique simultaneously guarantees threat relevance, technical correctness, pedagogical alignment, reproducibility, adaptability, and deployability. Future CTR scenario-generation systems should therefore combine complementary mechanisms: CTI and forecasting for threat relevance, structured scenario representations for consistency, generative AI for content diversity, RAG for external knowledge grounding, adaptive agents for runtime behaviour, and automated validation for deployment safety. Human oversight should remain part of this lifecycle to ensure that generated scenarios remain realistic, executable, and aligned with their intended training objectives (Zacharis et al. 2024; Braghin et al. 2019; Santorsola et al. 2022; Lupinacci et al. 2025).

A further cross-cutting lesson is that future AI-supported CTRs need to consider two complementary classes of attack scenarios. MITRE ATT&CK provides a structured basis for modelling adversarial behaviours targeting traditional ICT systems, whereas MITRE ATLAS extends adversarial knowledge representation towards attacks targeting AI and ML systems, including emerging LLM-based applications (MITRE 2026; Sachpelidis-Brozos et al. 2026). Integrating these knowledge bases into scenario-generation workflows could therefore support CTRs that train participants against both conventional cyber threats and AI-specific adversarial threats.

These lessons collectively address the research gaps summarised in Table 6 and provide the basis for the future research directions discussed in the section on future research directions.

Table 6: Research Gaps, Open Challenges, and Future Research Directions for AI-Based Scenario Generation in Cyber Training Ranges

Research Gap and Representative References	**Current State and Unresolved Challenge**	**Future Research Direction**
1. Fragmented scenario lifecycle *Representative references:* (Yamin et al. 2020; Fu et al. 2023; Lupinacci et al. 2025; Garg et al. 2025)	Existing approaches address scenario modelling, threat intelligence, orchestration, deployment, adaptation, and assessment through different tools and mechanisms, with limited integration across the complete scenario lifecycle.	Investigate integrated AI-assisted pipelines connecting scenario authoring, validation, deployment, execution, adaptation, assessment, and continuous improvement.
2. Formal semantics and verification *Representative references:* (Arshad et al. 2021; Braghin et al. 2019; Shirmohammadi and Tork Ladani 2025; Tomeschek et al. 2026; Lupinacci et al. 2025)	DSLs, model-driven approaches, and formal models provide structured scenario descriptions and support different forms of constraint checking and validation. However, validation of dependencies, attack paths, generated configurations, and learning objectives is not consistently integrated across scenario-generation approaches.	Integrate formal semantics, constraint checking, dependency validation, and automated generate–validate–repair mechanisms into AI-generated scenarios.
3. CTI integration *Representative references:* (Zacharis et al. 2024; Yamin et al. 2024; Palumickas et al. 2026)	Cyber Threat Intelligence (CTI) can improve scenario relevance, but threat information may be incomplete, outdated, noisy, or difficult to transform into coherent and executable training scenarios.	Develop trustworthy CTI-to-scenario pipelines using provenance tracking, relevance filtering, confidence assessment, RAG, and expert validation.
4. Reproducibility and benchmarking *Representative references:* (Palumickas et al. 2026; Lupinacci et al. 2025; Pavelić et al. 2025; Érsok et al. 2026)	Generative and agentic approaches can produce diverse outputs across repeated runs, while common datasets, prompts, configurations, and benchmarks for comparing scenario-generation systems remain limited.	Define reproducible evaluation protocols including shared benchmark scenarios, datasets, prompts, model configurations, retrieved context, and reporting criteria.
5. Behavioural realism *Representative references:* (Santorsola et al. 2022; Sánchez et al. 2023; Agrawal et al. 2026)	Many scenarios depend on predefined or scripted behaviour, while realistic modelling of changing attacker, defender, and user behaviour remains challenging.	Develop constrained, adaptive, multi-agent models that better represent realistic attacker, defender, and user behaviour while remaining aligned with scenario objectives and known Tactics, Techniques, and Procedures.

Continued on next page

Table 6: Research Gaps, Open Challenges, and Future Research Directions (continued)

Research Gap and Representative References	**Current State and Unresolved Challenge**	**Future Research Direction**
6. Learning-objective alignment *Representative references:* (Smyrlis et al. 2020; Chowdhury and Gkioulos 2023; Yamin et al. 2024; Gerontakis et al. 2023)	Existing approaches incorporate learning objectives, skills, competencies, and assessment criteria in different ways, but explicit linkage between generated scenario components and measurable learning objectives is not consistently represented across scenario-generation approaches.	Develop mechanisms that explicitly map generated events, tasks, attack steps, and scenario branches to Terminal Learning Objectives (TLOs), competencies, and assessment criteria.
7. Adaptive training *Representative references:* (Arshad et al. 2021; Yamin et al. 2022; Santorsola et al. 2022; Agrawal et al. 2026; Gizelis et al. 2025)	Existing approaches demonstrate performance-aware scenario selection and adaptive agents, but comprehensive real-time personalisation of scenario difficulty and progression remains limited.	Combine learner analytics, performance measures, reinforcement learning, and adaptive scenario logic to personalise difficulty, attack behaviour, and exercise progression.
8. Interoperability and portability *Representative references:* (Fu et al. 2023; Steininger 2025; Yamin and Katt 2025; Lal et al. 2025)	Scenario descriptions, orchestration mechanisms, and deployment artefacts are often dependent on particular cyber-range platforms, technologies, or configuration formats.	Develop platform-independent scenario representations, translation mechanisms, standard interfaces, and reusable orchestration models for cross-range and federated deployment.
9. Automated deployment *Representative references:* (Pham et al. 2016; Braghin et al. 2019; Bernardinetti et al. 2021; Caturano et al. 2022; Lupinacci et al. 2025)	Cyber-range instantiation and infrastructure deployment are increasingly automated, but deployment approaches remain dependent on platform-specific technologies, configuration requirements, and knowledge of the target cyber-range framework.	Develop more portable and reliable deployment pipelines by integrating Infrastructure-as-Code, automated configuration validation, testing, and framework-aware deployment mechanisms.
10. Scalability *Representative references:* (Gustafsson and Almroth 2020; Nakata and Otsuka 2021; Fu et al. 2023; Lal et al. 2025)	Large and complex scenarios increase infrastructure, orchestration, storage, and computational requirements, particularly when multiple virtual instances, users, scenarios, or federated cyber ranges must operate concurrently.	Investigate resource-efficient virtualisation, scalable and distributed orchestration, efficient resource allocation, and federation mechanisms that support increasing numbers of concurrent scenarios, users, and cyber-range infrastructures.

Continued on next page

Table 6: Research Gaps, Open Challenges, and Future Research Directions (continued)

Research Gap and Representative References	Current State and Unresolved Challenge	Future Research Direction
11. Evaluation of scenario quality *Representative references:* (Palumickas et al. 2026; Lupinacci et al. 2025; Pavelić et al. 2025; Érsok et al. 2026)	Existing studies evaluate generated scenarios using different technical, realism, diversity, deployment, and training-related criteria, but evaluation remains fragmented across studies and scenario-generation approaches.	Develop multidimensional evaluation frameworks that combine technical correctness, realism, diversity, deployability, and educational effectiveness, supported by automated testing and expert-based validation.
12. Data and knowledge quality *Representative references:* (Zacharis et al. 2024; Vadisetty and Polamarasetti 2024; Palumickas et al. 2026)	AI-assisted scenario generation depends on the relevance, freshness, representativeness, and quality of training data, CTI, and retrieved knowledge; increasing the volume of available information does not necessarily improve generated scenario quality.	Develop provenance-aware knowledge sources, cybersecurity-specific retrieval and ranking, quality filtering, data validation, and mechanisms for maintaining current threat knowledge.
13. Explainability, safety, and human oversight *Representative references:* (Yamin et al. 2024; Reuter et al. 2025; Pavelić et al. 2025; Érsok et al. 2026; Lupinacci et al. 2025)	LLMs and increasingly autonomous agents can introduce hallucinations, technical inconsistencies, unsupported actions, and limited transparency, creating additional reliability and safety challenges for AI-assisted scenario generation.	Develop explainable and policy-aware generation, capability constraints, human-in-the-loop approval, provenance tracking, and automated safety and technical validation.
14. Accessibility and authoring complexity *Representative references:* (Reuter et al. 2025; Rizos et al. 2025; Lupinacci et al. 2025)	Developing and configuring technically realistic cyber-range scenarios still requires substantial cybersecurity and platform-specific expertise, although recent LLM-based approaches reduce some of the authoring burden.	Develop natural-language authoring interfaces, reusable templates, AI-assisted design tools, and framework-aware generation that enables instructors to create scenarios without mastering platform-specific configuration syntax.

Note: Representative references provide selected examples of relevant studies identified in the literature review and are not intended to be exhaustive. Research gaps and future directions reflect the authors' synthesis.

7 Research Challenges in AI-Based Scenario Generation

Despite recent advances, several research challenges remain before AI-based scenario generation can support dependable, scalable, and autonomous CTR operation. The literature identifies challenges associated with formal scenario validation, data quality, reproducibility, technical realism, computational scalability, interoperability, pedagogical alignment, and trustworthy AI. These challenges become increasingly important as scenario generation progresses from conventional ML and DL towards generative models, LLMs, RAG, and agentic systems. The following research directions are derived from the lessons learned presented in Section 6.

1) Cyber Training Range Scenario Generation Based on MITRE ATT&CK and MITRE ATLAS-Based Attacks.

A key research challenge is developing AI-supported CTR scenario-generation methods that can address two complementary target types: traditional ICT systems and AI-based systems, including ML models and LLM-based applications. For traditional ICT environments, MITRE ATT&CK provides a structured representation of adversary tactics and techniques that can support the generation of threat-informed attack scenarios targeting networks, hosts, applications, and other conventional computing infrastructure. In contrast, MITRE ATLAS provides a complementary knowledge base for adversarial threats targeting AI and ML systems, including emerging attacks against LLM-based applications (MITRE 2026; Sachpelidis-Brozos et al. 2026). Future research should investigate how ATT&CK- and ATLAS-aligned attack knowledge can be incorporated into AI-supported scenario-generation pipelines to automatically generate, validate, and deploy realistic scenarios for both conventional ICT and AI-enabled environments. This challenge is directly related to the need for knowledge-grounded and technically reliable scenario generation identified in Lessons 4 and 8.

2) Formal Semantics and Automated Scenario Validation.

A fundamental research challenge concerns the formal representation and validation of automatically generated scenarios, as identified in Lessons 4, 6, and 7. Existing model-driven approaches demonstrate that structured scenario descriptions and formal models can be used to identify inconsistencies before deployment (Braghin et al. 2019; Russo et al. 2020). However, AI-generated scenarios introduce additional uncertainty because individual components may appear plausible while being collectively inconsistent. For example, a generated attack sequence may require a vulnerability that is absent from the target system, an event may be triggered before its prerequisite has occurred, or an attack path may reach a state from which the intended training objective cannot be completed.

This challenge becomes particularly important for LLM- and agent-based generation because the scenario is no longer completely predetermined by a human designer. Future research should therefore investigate machine-readable scenario representations with explicit structural, semantic, and operational constraints. Validation mechanisms should determine whether the required infrastructure can be instantiated, whether dependencies and attack paths are technically feasible, whether event sequences are

logically consistent, and whether the intended TLOs can be reached (Braghin et al. 2019; Russo et al. 2020; Lupinacci et al. 2025).

A promising research direction is the integration of generative AI with formal or constraint-based validation through a *generate–validate–repair* process. In this approach, AI first generates the scenario, an independent validation mechanism identifies structural or semantic violations, and the generation system then repairs the identified problems before deployment. This would combine the flexibility of AI generation with the reliability of formally constrained scenario engineering.

3) Reliable, Current, and Context-Specific Cybersecurity Data.

AI-based scenario generation depends strongly on the quality and relevance of the information used for model training, threat forecasting, retrieval, and scenario grounding, as identified in Lessons 1 and 4. Cybersecurity data can be incomplete, outdated, noisy, or biased towards particular technologies, sectors, or attack types. Consequently, an AI model may generate a technically plausible scenario that does not accurately represent the current threat environment of the organisation being trained (Al Siam et al. 2025; Zacharis et al. 2024).

The evaluation of ASCERT further demonstrates that increasing the amount of information supplied to an AI system does not necessarily improve scenario quality. Focused and carefully selected knowledge produced better correspondence with real incidents than larger collections containing less relevant information (Palumickas et al. 2026). Similarly, generative models such as GANs and VAEs depend on sufficiently representative and diverse cybersecurity data to produce realistic attack variations (Vadisetty and Polamarasetti 2024).

Future research should therefore focus not only on collecting larger datasets, but also on assessing data quality, freshness, relevance, diversity, and provenance. For RAG-based systems, cybersecurity-specific retrieval and ranking mechanisms are required to select information relevant to a particular organisation, sector, technology stack, threat actor, and training objective. Continuous integration of validated CTI represents a promising approach to maintaining scenario relevance as the threat landscape evolves.

4) Reproducibility, Difficulty Control, and Fair Assessment.

The non-deterministic behaviour of generative AI creates an important tension between scenario diversity and assessment consistency, reflecting the issues identified in Lessons 2 and 6. Dynamic generation is valuable because trainees cannot simply memorise a predetermined attack sequence. However, the same or similar inputs may produce different attack paths, response choices, and scenario outcomes. The ASCERT evaluation, for example, demonstrates substantial variation between scenario runs (Palumickas et al. 2026).

This variability becomes particularly important when CTRs are used for formal assessment or competency measurement. Two trainees may receive scenarios that differ in technical complexity, number of attack stages, available response options, or overall difficulty. Their resulting performance scores may therefore not be directly comparable.

Future research should investigate *controlled variability*, in which scenarios remain sufficiently diverse to prevent memorisation while preserving equivalent TLOs and comparable levels of difficulty. This requires research into difficulty calibration, scenario

equivalence, reproducibility controls, and standardised performance metrics. Learning analytics should also extend beyond individual exercises to support longitudinal modelling of trainee progress and competency development (Grimaldi et al. 2023).

5) Technical Realism and Executable Scenario Generation.

Another major challenge is closing the gap between generating a plausible cybersecurity description and generating a technically executable CTR scenario, as highlighted particularly in Lessons 4 and 6. An AI system may correctly describe ransomware, credential theft, lateral movement, or data exfiltration while failing to produce the infrastructure, vulnerabilities, commands, dependencies, network traffic, or attack artefacts required to execute the scenario within a CTR (Garg et al. 2025; Hannay et al. 2021).

This distinction is particularly important for LLM-based generation. Natural-language plausibility does not guarantee that generated commands, configuration files, Internet Protocol addressing, software versions, vulnerabilities, or attack sequences are technically valid. The ASCERT evaluation, for example, identifies problems involving attack-stage ordering and generated scenario actions (Palumickas et al. 2026). Similarly, ARCeR identifies the challenge of determining whether a requested scenario is actually supported by the capabilities of the target CTR platform (Lupinacci et al. 2025).

Future research should therefore progress from *scenario-description generation* towards *executable scenario generation*. Generated scenarios should be automatically tested in isolated environments before being released to trainees. Such validation could include syntax checking, dependency validation, topology verification, vulnerability verification, attack-path execution, and automated rollback when deployment fails.

Technical realism should also extend beyond isolated IT systems. Cybersecurity incidents involving critical infrastructure may propagate between interconnected cyber and physical systems. Future scenario-generation approaches should therefore improve their ability to represent cascading effects in which compromise of one component affects other services or operational processes (Karjalainen and Kokkonen 2020).

6) Scalability and Computational Cost.

Increasing the sophistication of AI-based scenario generation also increases computational requirements, particularly as systems incorporate adaptive and agentic behaviour as discussed in Lessons 3 and 5. DL models, RL environments, multi-agent simulations, LLMs, and RAG pipelines can require substantial processing, memory, storage, and networking resources. Adaptive multi-agent environments can further increase this demand because multiple attacker and defender agents may operate while the underlying CTR infrastructure is simultaneously executing (Agrawal et al. 2026).

This creates an important scalability challenge. An approach that performs successfully in an experimental environment may not remain practical when many trainees, teams, scenarios, and AI agents operate concurrently. This is particularly relevant to educational institutions, small organisations, and portable or resource-constrained CTR deployments.

Future research should therefore investigate resource-efficient models, scenario caching, model optimisation, distributed processing, and hybrid cloud architectures in which computationally expensive AI operations are invoked only when required. Evaluation of AI-based scenario generation should consequently consider not only

scenario quality but also generation latency, computational cost, resource utilisation, and scalability with increasing numbers of concurrent trainees and scenarios.

7) Interoperability, Portability, and CTR Federation.

AI-generated scenarios will have limited reuse if they remain tightly coupled to a particular CTR platform. This issue is closely related to the need for integrated and deployable scenario-generation architectures identified in Lessons 5 and 8. Existing CTRs use different virtualisation technologies, topology descriptions, scenario formats, orchestration mechanisms, Application Programming Interfaces (APIs), and assessment systems. Consequently, a scenario generated for one platform may require substantial modification before it can be executed on another (Davis and Magrath 2013; Kampourakis et al. 2025).

This challenge becomes increasingly important for federated CTRs, where multiple geographically or administratively independent infrastructures may participate in a common exercise. AI-generated scenario components must therefore describe not only attack behaviour but also infrastructure requirements and dependencies in a portable and interoperable form.

Future research should investigate common scenario schemas, interoperable APIs, abstraction layers, and platform-independent intermediate representations that can subsequently be translated into platform-specific configurations. Agentic approaches such as ARCeR provide an important direction because external documentation can guide configuration generation for different CTR frameworks (Lupinacci et al. 2025). However, automated capability discovery is still required so that an AI system can determine whether the target platform supports the requested scenario before attempting deployment.

The related challenge of reproducing realistic organisational environments also remains significant. High-fidelity replicas of enterprise and critical-infrastructure environments can require substantial computational and configuration resources (Orbinato 2021; Priyadarshini 2018). Future research should therefore investigate methods for automatically determining the level of infrastructure fidelity required to satisfy the intended training objectives without unnecessarily reproducing every component of the real environment.

8) Pedagogical Alignment and Competency-Aware Generation.

A technically realistic scenario is not necessarily an effective training scenario. This challenge is closely related to the requirements for adaptive training and human-guided educational alignment identified in Lessons 3 and 7. Existing CTR automation research frequently concentrates on infrastructure, attack execution, or scenario events, while the relationship between automatically generated activities and measurable cybersecurity competencies requires further investigation (Pramon et al. 2025; Hannay et al. 2021).

Future AI-based scenario generation should therefore begin not only with a technical description of an attack but also with explicit TLOs. The AI system should be able to translate these objectives into appropriate scenario events, trainee tasks, expected responses, difficulty levels, and measurable assessment criteria.

An important research direction is *competency-aware scenario generation*, in which generated content is explicitly mapped to recognised cybersecurity knowledge, skills, and competency frameworks. This would allow an instructor to request a scenario

designed to develop or assess particular competencies rather than merely specifying an attack type.

Research is also required to represent team-based performance. Cybersecurity incidents are normally handled collaboratively, and future systems should therefore distinguish individual technical competence from team-level coordination, communication, leadership, and decision-making (Yamin et al. 2020).

9) Explainability, Safety, and Dual-Use Risk.

Greater AI autonomy creates important safety and governance challenges, particularly in relation to the human oversight requirements identified in Lessons 7 and 8. AI-generated cybersecurity knowledge has an inherent dual-use characteristic: techniques that enable defenders to train against sophisticated attacks may also provide information that can support malicious activity (Kaloudi and Li 2020; Yamin et al. 2024). This concern becomes increasingly important as systems progress from describing attacks towards generating attack artefacts, infrastructure configurations, or autonomous adversarial behaviour.

Explainability presents a related challenge. Complex AI systems may select an attack path or modify a scenario without providing sufficient information about why that decision was made. Limited transparency makes it difficult for instructors to determine whether generated scenario behaviour is technically appropriate, pedagogically justified, or suitable for assessment (Lillemets et al. 2025).

Future research should therefore investigate explainable scenario-generation mechanisms that record the evidence, retrieved knowledge, constraints, and generation decisions associated with important scenario components. Generated content should also be subject to appropriate security controls, sandboxing, access restrictions, audit logging, and human approval according to its potential risk.

Agentic systems introduce an additional challenge because an AI agent can coordinate multiple operations and interact with external tools and infrastructure. ARCeR demonstrates the potential of agentic automation but also identifies the need for improved capability validation and semantic correction (Lupinacci et al. 2025). Future systems should therefore avoid depending solely on the generating AI model to determine whether its own output is technically valid or safe. Independent validation mechanisms and human oversight should remain part of the scenario-generation pipeline.

10) Balancing AI Autonomy with Human Control.

The combined challenges above indicate that the central research problem is not simply how to maximise automation, but how to determine the appropriate division of responsibility between AI and human experts. This challenge brings together the issues of agentic autonomy and human oversight identified particularly in Lessons 5 and 7. Earlier CTR automation approaches demonstrate the value of automated instantiation, orchestration, and scenario randomisation while maintaining explicit control over important exercise structures (Gustafsson and Almroth 2020; Nakata and Otsuka 2021). AI-based systems introduce greater flexibility and adaptability but also greater uncertainty.

A promising research direction is therefore a *human-in-the-loop* scenario-generation architecture. AI can perform computationally intensive and repetitive activities such as

CTI analysis, threat forecasting, scenario variation, configuration generation, and adaptive event selection, while human instructors retain authority over TLOs, acceptable risk, technical approval, and final deployment.

Ultimately, the next generation of AI-enabled CTRs will require more than increasingly powerful generative models. A dependable scenario-generation system must combine AI generation with formal scenario representations, high-quality CTI, technical validation, controlled variability, competency mapping, interoperability, explainability, and human governance. Addressing these research challenges is necessary to progress AI-based scenario generation from experimental prototypes towards trustworthy, scalable, technically executable, and pedagogically defensible cybersecurity training systems.

8 Conclusion

This systematic literature review traces the evolution of scenario generation in CTRs and highlights the growing impact of AI. It shows a shift from traditional methods, such as manually crafted, template-based, rule-based, attack-graph-based, and model-driven techniques, towards advanced AI-supported approaches. These newer methods include ML, DL, generative models, RL, LLMs, RAG, and emerging agentic AI. Although traditional non-AI approaches offer benefits such as structure, repeatability, formalisation, and control over deployment, AI enhances scenario generation by increasing diversity, threat relevance, adaptability, and automation. The review demonstrates that various AI techniques play distinct roles at different stages of scenario generation, rather than offering a complete solution on their own. ML and DL assist with threat analysis and prediction; generative models enhance the variety of cyber-threat artifacts; and RL, along with multi-agent approaches, enables adaptive adversarial behaviours. LLMs expand these abilities by creating multiple linked scenario components from natural-language inputs, while RAG grounds the generated content in external cybersecurity knowledge. Emerging agentic methods further this development through reasoning, retrieval, tool utilisation, iterative validation, configuration creation, and deployment. Consequently, AI-generated and adaptive attack scenarios are a key, cross-cutting capability that can be achieved through several of these technologies rather than as a standalone scenario-generation method. However, advancing AI capabilities alone does not guarantee reliable or educationally effective CTR scenarios. The review highlights ongoing issues related to data quality and timeliness, hallucinations, technical inconsistencies, formal scenario validation, reproducibility, difficulty management, interoperability, portability, computational scalability, alignment with learning objectives, explainability, safety, and scenario quality evaluation. Notably, creating a plausible cybersecurity scenario differs fundamentally from developing a technically executable scenario that can be safely deployed within a CTR and reliably used for training or assessment. Future research on CTR scenario generation should go beyond simply increasing AI autonomy and instead emphasise trustworthy hybrid systems and human-in-the-loop designs. These systems should integrate AI-driven generation and adaptation with high-quality CTI, structured scenario models, formal validation, controlled variability, competency-based learning goals, interoperable deployment, explainability, and human oversight. Meeting these criteria can help create

CTRs that are more automated and adaptable, while also being realistic, reproducible, technically feasible, pedagogically sound, and reliable for cybersecurity training. Even though this survey focuses on CTR scenario generation, it has a broader implication. The insights gained from this survey will benefit relevant research communities working on scenario generation for digital twins, airline simulation, intelligent transportation systems, robotic navigation, disaster recovery systems, and so on.

Statements and Declarations

Funding The authors acknowledge support from the Internet Commerce Security Lab at Federation University. Purvi Mehta acknowledges support through an Australian Government Research Training Program (RTP) Fee-Offset Scholarship at Federation University, Australia.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this manuscript, Generative AI is used for language editing, including grammar correction and improvement of readability.

References

Agrawal A, Nadeem M, Al Nuaim A, et al (2026) Artificial intelligence driven multi agent framework for adaptive cyber attack simulation and automated incident response in cyber range environments. Scientific Reports

Al Siam A, Alazab M, Awajan A, et al (2025) A comprehensive review of ai's current impact and future prospects in cybersecurity. IEEE Access 13:14029–14050

Arshad S, Alam M, Al-Kuwari S, et al (2021) Attack specification language: Domain specific language for dynamic training in cyber range. In: 2021 IEEE Global Engineering Education Conference (EDUCON), IEEE, pp 873–879

Bernardinetti G, Iafrate S, Bianchi G (2021) Nautilus: A tool for automated deployment and sharing of cyber range scenarios. In: Proceedings of the 16th International Conference on Availability, Reliability and Security, pp 1–7

Beuran R, Inoue T, Tan Y, et al (2019) Realistic cybersecurity training via scenario progression management. In: 2019 IEEE European symposium on security and privacy workshops (EuroS&PW), IEEE, pp 67–76

Braghin C, Cimato S, Damiani E, et al (2019) A model driven approach for cyber security scenarios deployment. In: International Workshop on Information and Operational Technology Security Systems, Springer, pp 107–122

Caturano F, d'Ambrosio N, Perrone G, et al (2022) Exploitwp2docker: a platform for automating the generation of vulnerable wordpress environments for cyber ranges. In: 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), IEEE, pp 1–7

Chouliaras N, Kittes G, Kantzavelou I, et al (2021) Cyber ranges and testbeds for education, training, and research. Applied Sciences 11(4):1809

Chowdhury N, Gkioulos V (2023) A personalized learning theory-based cyber-security training exercise. International Journal of Information Security 22(6):1531–1546

Costa G, Russo E, Armando A (2022) Automating the generation of cyber range virtual scenarios with VSDL. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications 13(4):61–80. https://doi.org/10.58346/JOWUA.2022.I4.004

Dalla Costa A, Kuusijärvi J (2022) Programmatic description language for cyber range topology creation. In: 2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), IEEE, pp 403–412

Davis J, Magrath S (2013) A survey of cyber ranges and testbeds. General Document DSTO-GD-0771, Defence Science and Technology Organisation, Edinburgh, South Australia, URL https://apps.dtic.mil/sti/pdfs/ADA594524.pdf

Érsok M, Bánáti A, Kail E, et al (2026) Agentic language models for vulnerability-based cyber range scenario generation. In: 2026 49th MIPRO ICT and Electronics Convention (MIPRO), IEEE, pp 771–775

Fu Y, Han W, Yuan D (2023) Disentangled orchestration on cyber ranges. IEEE Transactions on Dependable and Secure Computing 21(4):2344–2360

Garg A, Boualouache A, Imeri A, et al (2025) A survey of cyber range training exercise scenario description, generation, and execution. Authorea Preprints

Gerontakis G, Yannakopoulos P, Voyiatzis I (2023) Evaluating cybersecurity certifications: A framework for extracting educational scenarios in cybersecurity training. In: Proceedings of the 27th Pan-Hellenic Conference on Progress in Computing and Informatics, pp 243–248

Gizelis CA, Nikoloudakis N, Papanikas D, et al (2025) Developing adaptive ai cyber range training modules for critical infrastructure sectors. In: IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, pp 157–167

Glas M, Vielberth M, Pernul G (2023) Train as you fight: evaluating authentic cyber-security training in cyber ranges. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp 1–19

Grimaldi A, Ribiollet J, Nespoli P, et al (2023) Toward next-generation cyber range: A comparative study of training platforms. In: European Symposium on Research in Computer Security, Springer, pp 271–290

Gustafsson T, Almroth J (2020) Cyber range automation overview with a case study of crate. In: Nordic Conference on Secure IT Systems, Springer, pp 192–209

Hannay JE, Stolpe A, Yamin MM (2021) Toward ai-based scenario management for cyber range training. In: International Conference on Human-Computer Interaction, Springer, pp 423–436

Kaloudi N, Li J (2020) The ai-based cyber threat landscape: A survey. ACM Computing Surveys (CSUR) 53(1):1–34

Kampourakis V, Gkioulos V, Katsikas S (2025) A step-by-step definition of a reference architecture for cyber ranges. Journal of Information Security and Applications 88:103917

Karjalainen M, Kokkonen T (2020) Comprehensive cyber arena; the next generation cyber range. In: 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), IEEE, pp 11–16

Lal C, Grammatopoulos AV, Takaronis M, et al (2025) Cyberunity: a federated architecture for next-generation cybersecurity training. In: 2025 IEEE International Conference on Cyber Security and Resilience (CSR), IEEE, pp 759–764

Leitner M, Frank M, Hotwagner W, et al (2020) Ait cyber range: flexible cyber security environment for exercises, training and research. In: Proceedings of the 2020 European Interdisciplinary Cybersecurity Conference, pp 1–6

Lillemets P, Bashir Jawad N, Kashi J, et al (2025) A systematic review of cyber range taxonomies: Trends, gaps, and a proposed taxonomy. Future Internet 17(6):259

Lupinacci M, Blefari F, Romeo F, et al (2025) Arcer: an agentic rag for the automated definition of cyber ranges. In: International Conference on Availability, Reliability and Security, Springer, pp 23–40

MITRE (2026) MITRE ATT&CK. https://www.mitre.org/focus-areas/cybersecurity/mitre-attack, accessed: 2026-08-26

Nakata R, Otsuka A (2021) Cyexec*: A high-performance container-based cyber range with scenario randomization. IEEE Access 9:109095–109114

National Institute of Standards and Technology (2023) The Cyber Range: A Guide. URL https://www.nist.gov/system/files/documents/2023/09/29/The%20Cyber%20Range_A%20Guide.pdf, accessed: 19 October 2026

Orbinato V (2021) A next-generation platform for cyber range-as-a-service. In: 2021 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), IEEE, pp 314–318

Page MJ, McKenzie JE, Bossuyt PM, et al (2021) The prisma 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372:n71. https://doi.org/10.1136/bmj.n71

Palumickas M, Yamin MM, Katt B, et al (2026) Evaluating ascert: generative ai for cyber-range scenario generation. International Journal of Information Security 25(1):8

Pavelić D, Kovačević I, Katulić F, et al (2025) Towards agent-based generation of multi-step vulnerability scenarios for cyber ranges. In: 2025 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), IEEE, pp 1–6

Pfaller T, Skopik F, Smith P, et al (2025) Exploring a comprehensive approach to customize cyber exercises utilizing a process-based lifecycle model. International Journal of Information Security 24(2):96

Pham C, Tang D, Chinen Ki, et al (2016) Cyris: A cyber range instantiation system for facilitating security training. In: Proceedings of the 7th Symposium on Information and Communication Technology, pp 251–258

Pramon N, Wannapiroon P, Nilsook P (2025) Conceptual framework of artificial intelligence strategy with scenario-based training. In: 2025 5th International Conference on Educational Communications and Technology (ICTAECT), IEEE, pp 1–8

Priyadarshini I (2018) Features and architecture of the modern cyber range: a qualitative analysis and survey. University of Delaware

Reuter L, Smith P, Skopik F (2025) Ai-augmented scenario design: Experiences from a national cybersecurity exercise. In: International Conference on Critical Information Infrastructures Security, Springer, pp 305–325

Ribaudo M, Valenza A (2019) Semi-automatic generation of cybersecurity exercises: A preliminary proposal. In: Proceedings of the 2nd ACM SIGSOFT International Workshop on Ensemble-Based Software Engineering for Modern Computing Platforms, pp 16–21

Rizos GS, Kopalidis N, Mengidis N, et al (2025) From concept to deployment: An ai assistant for generating and configuring cyber range scenarios. In: 2025 IEEE International Conference on Cyber Security and Resilience (CSR), IEEE, pp 777–782

Russo E, Costa G, Armando A (2018) Scenario design and validation for next generation cyber ranges. In: 2018 IEEE 17th International Symposium on Network Computing and Applications (NCA), IEEE, pp 1–4

Russo E, Costa G, Armando A (2020) Building next generation cyber ranges with crack. Computers & Security 95:101837

Sachpelidis-Brozos N, Katsaros E, Radoglou-Grammatikis P, et al (2026) Hacking intelligence: A systematic review of adversarial attacks on artificial intelligence through the mitre atlas framework. Computer Science Review 61:100923

Sánchez PM, Nespoli P, Alfaro JG, et al (2023) Methodology for automating attacking agents in cyber range training platforms. In: International Workshop on Secure and Resilient Digital Transformation of Healthcare, Springer, pp 90–109

Santorsola A, Migliau A, Caporusso S (2022) Reinforcement learning agents for simulating normal and malicious actions in cyber range scenarios. In: ITASEC, pp 1–16

Shirmohammadi N, Tork Ladani B (2025) CST-SDL: A scenario description language for collaborative security training in cyber ranges. ISC International Journal of Information Security 17(1):59–73. https://doi.org/10.22042/ISECURE.2024.462008.1132

Smyrlis M, Fysarakis K, Spanoudakis G, et al (2020) Cyber range training programme specification through cyber threat and training preparation models. In: International Workshop on Model-Driven Simulation and Training Environments for Cybersecurity, Springer, pp 22–37

Somarakis I, Smyrlis M, Fysarakis K, et al (2019) Model-driven cyber range training: a cyber security assurance perspective. In: International workshop on information and operational technology security systems, Springer, pp 172–184

Stamatopoulos D, Katsantonis M, Fouliras P, et al (2024) Exploring the architectural composition of cyber ranges: A systematic review. Future Internet 16(7):231

Steininger C (2025) Creating a framework for platform-independent cyber range scenarios. In: NOMS 2025-2025 IEEE Network Operations and Management Symposium, IEEE, pp 1–4

Steininger C, Götz L, Schopp M (2025) How accessible is cybersecurity training? a survey on the accessibility, capabilities, and technology stack of cyber ranges. IEEE Access 13:203980–204039. https://doi.org/10.1109/ACCESS.2025.3638168

Tomeschek S, Jungbauer C, Luidold C (2026) A design science approach to modular scenario modeling for cyber security exercises. In: IFIP World Conference on Information Security Education, Springer, pp 130–144

Ukwandu E, Farah MAB, Hindy H, et al (2020) A review of cyber-ranges and test-beds: Current and future trends. Sensors 20(24):7148

Vadisetty R, Polamarasetti A (2024) Generative ai for cyber threat simulation and defense. In: 2024 12th International Conference on Control, Mechatronics and Automation (ICCMA), IEEE, pp 272–279

Yamin MM, Katt B (2022) Modeling and executing cyber security exercise scenarios in cyber ranges. Computers & Security 116:102635

Yamin MM, Katt B (2025) Modelling attack and defense scenarios on federated cyber ranges. In: 2025 IEEE International Conference on Cyber Security and Resilience (CSR), IEEE, pp 771–776

Yamin MM, Katt B, Gkioulos V (2020) Cyber ranges and security testbeds: Scenarios, functions, tools and architecture. Computers & Security 88:101636

Yamin MM, Shukla A, Ullah M, et al (2022) Adapt-automated defence training platform in a cyber range. In: International Conference on Information Systems and Management Science, Springer, pp 184–203

Yamin MM, Hashmi E, Ullah M, et al (2024) Applications of llms for generating cyber security exercise scenarios. IEEE Access

Zacharis A, Katos V, Patsakis C (2024) Integrating ai-driven threat intelligence and forecasting in the cyber security exercise content generation lifecycle. International Journal of Information Security 23(4):2691–2710

Zhang J, Li Z, Chen R, et al (2023) A survey of cyber range: Current status, analysis, and future trends. In: International Conference on Network Simulation and Evaluation, Springer, pp 88–101