Data-Driven Innovation in Cybersecurity for E-Business Information Systems: A Case Study on CVE Report Generation via LLM[*]

Arcangelo Castiglione, Francesco Colace, Marco Fusco,
Antonio Giordano, Domenico Santaniello, Carmine Valentino[†]

University of Salerno, Fisciano (SA), Italy
{arcastiglione, fcolace, dsantaniello, cvalentino}@unisa.it,
{m.fusco37, a.giordano223}@studenti.unisa.it

Abstract

The backbone of modern digital economies, e-business information systems (EBIS), is security, which is an essential factor in ensuring business continuity and digital trust. One of the most complex challenges in this area is efficiently managing vulnerabilities, which requires fast and accurate reporting processes.

This study explores the application of artificial intelligence as* data-driven innovation to automate the generation of structured cybersecurity reports. To this end, we analyzed the performance of four open-source large language models (LLMs) by implementing an adaptive two-step prompt chaining architecture. The system, tested on a dataset of 30 Common Vulnerabilities and Exposures (CVEs), simulates a realistic analysis and review workflow.

The assessment, based on quantitative and qualitative metrics, produced results that, although preliminary, confirm the feasibility of integrating LLMs into EBIS cybersecurity processes. The study also highlights crucial challenges related to scalability and reliability, which are key determinants of the governance of these systems.

1 Introduction

In the digital age, E-Business Information Systems (EBIS) form the backbone of modern economies, enabling everything from decentralized commerce (Web 3.0) to intelligent service infrastructures[1]. Their increasing complexity and interconnectedness, especially with the advent of 6G and ultra-connected ecosystems, expose them to an ever-changing cyber threat landscape. Efficient vulnerability management, cataloged in the Common Vulnerabilities and Exposures (CVE) database, is not just a technical function but a strategic pillar to ensure operational resilience and digital trust[2], [3].

This environment requires urgent business process innovation based on data and artificial intelligence. Security Operations Centers (SOCs), which are responsible for defending EBIS, are overwhelmed by a volume of data that cannot be manually managed. Large Language Models (LLMs)

[*] Proceedings of The 2025 IFIP WG 8.4 International Symposium on E-Business Information Systems Evolution (EBISION 2025), Article No. 5, December 16-18, 2025, Sapporo, Japan. © The copyright of this paper remains with the author(s).

[†] Corresponding author

emerge as a transformative technology for automating complex cognitive tasks like security report generation[4], [5], [6].

In this scenario, the present work aims to evaluate the effectiveness of four state-of-the-art open-source LLMs experimentally. The novelty of the proposed research lies in the proposal and validation of an adaptive architecture based on two-phase prompt chaining, tested on a dataset of 30 CVEs. Unlike previous studies focused on code classification or repair, our approach simulates an end-to-end process of writing and reviewing. This makes it possible to evaluate not only the initial generation capacity of models, but also the ability to self-improve, a key factor in building reliable AI systems for industrial contexts.

The organization of this paper is as follows: Section 2 outlines the background on EBIS, CVE, and LLM. Section 3 analyzes related jobs. Section 4 describes the proposed solution architecture as a resilient system for EBIS. Section 5 details the practical implementation. Section 6 presents the experimental results and discusses the governance implications of EBISs. Finally, Section 7 outlines conclusions and future research paradigms.

2 Background

An E-Business Information System (EBIS) is a socio-technical infrastructure that leverages digital technologies to run, monitor, and manage online business processes. These systems encompass various applications, from e-commerce platforms to decentralized financial systems to cloud-native architectures[7]. Their security plays a crucial role: a single vulnerability can compromise sensitive data, disrupt operations, and undermine consumer trust, generating potentially devastating impacts on the business. In this context, the management of Common Vulnerabilities and Exposures (CVEs) is a strategic activity within an EBIS's governance to ensure its resilience against cyber attacks[8], [9].

We live today in the era of artificial intelligence, in which Large Language Models (LLMs) represent one of the most influential technologies[10], [11], [12]. These models, trained on vast textual datasets, are able to understand and generate natural language, offering new possibilities for automation and decision support. The models selected for this study reflect different design philosophies:

- DeepSeek-R1-Distill-Qwen-14B: A large model (14B parameters), known for its high-quality text output;
- Mistral-7B-Instruct-v0.1: an efficient model (7B parameters), optimized to follow complex instructions;
- WhiteRabbitNeo-13B: a 13B parameter model, designed for tasks that require high adherence to structured formats;
- Lily-7B-Instruct-v0.2: A 7B parameter model, optimized explicitly for cybersecurity, represents an example of specialized artificial intelligence.

The application of AI to cybersecurity is now a rapidly expanding research paradigm, with a significant impact on the evolution of EBIS. Numerous studies have explored the use of LLMs in this area, albeit with different approaches and purposes[13], [14], [15].

For example, the usefulness of language models in vulnerability analysis and assessment [16] has been demonstrated, although without particular attention to generating comprehensive reports for analysts. Other contributions focused on self-healing code [17], [18], highlighting a more governance-oriented and risk-communicating application[19].

Some studies have investigated the synthesis capacity of LLMs [20], while others have proposed chatbots for vulnerability analysis [21]. In general, the literature confirms the enormous potential of such models, but underlines the need for practical case studies that assess their reliability in realistic operational scenarios [22].

The present work is distinguished by the adoption of a two-step prompt chaining methodology, which allows to obtain a more in-depth evaluation of the ability of an LLM to operate within an iterative business workflow. This approach contributes in a concrete way to the debate on how to design more resilient, adaptive and intelligent EBIS architectures, capable of dynamically responding to emerging digital security challenges[23].

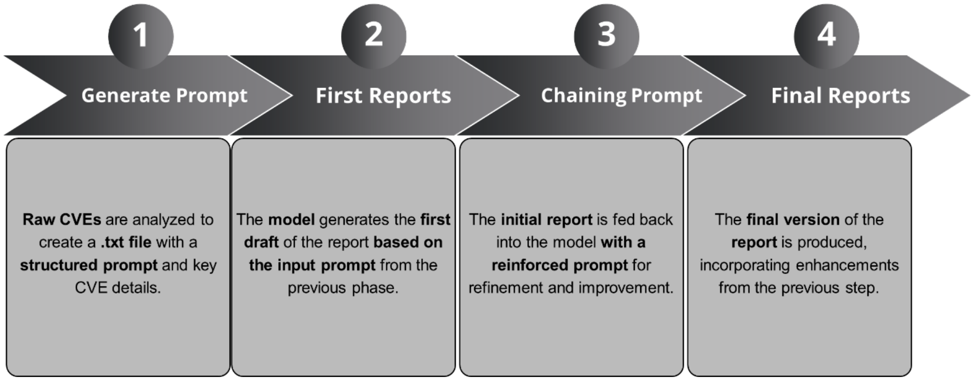
The proposed approach: An Adaptive System for Security Reporting

This section discusses the proposed approach, which introduces a system architecture for automating the security reporting lifecycle, designed to be integrated into a modern E-Business Information System (EBIS)[24].

The architecture is adaptive and iterative, as it incorporates a self-review mechanism based on prompt chaining, capable of progressively improving the quality of the output and reducing the impact of errors generated in the early stages of the process [24], [25].

The workflow is designed to simulate a hierarchical interaction between a junior analyst and a senior auditor, consisting of four automated phases, each of which represents a functional node of the system (Figure 1):

1. Initial Prompt Generation – The appropriately structured CVE data is converted into a detailed prompt for the "junior analyst." This prompt defines the structure of the report and constrains its format (sections: Summary, Technical Details, Risk Assessment, Suggested Mitigations, Exploitation Scenario).
2. First Report Production – The LLM generates a preliminary draft of the security report, adhering to the structural constraints of the initial prompt.
3. Generation of the review prompt – The report produced is incorporated into a second prompt that instructs the LLM to act as a "senior reviewer", with the aim of improving consistency, technical accuracy and clarity of presentation.
4. Production of the final report – The LLM refines the paper, correcting imperfections and completing missing sections. The result is a standardized and validated final report, ready to be distributed within the company information system.



Raw CVEs are analyzed to create a **.txt file** with a **structured prompt** and key CVE details.

The **model** generates the **first draft** of the report **based on the input prompt** from the previous phase.

The **initial report** is fed back into the model **with a reinforced prompt** for refinement and improvement.

The **final version** of the **report** is produced, incorporating enhancements from the previous step.

Figure 1. Methodological workflow based on prompt chaining

This architecture not only optimizes the efficiency of the security analysis and reporting process, but also introduces an automated quality control mechanism, which is crucial for the governance of information systems and for maintaining high levels of digital trust. It also allows you to monitor key indicators such as success rate, report completeness, average generation time and structural adherence, which are essential for the comparative evaluation of the performance of the language models used.

To complete the architectural description, Figure 2 shows the flow of data within the system, highlighting the input, processing, and output steps.

In this representation, the raw CVE data constitutes the process input ("RAW CVEs"), which are first integrated into the initial prompt and then transformed into an intermediate report. This report feeds a second round of LLM inference through the chain prompt, eventually producing a final report structured in text format.

This diagram (Figure 2) clearly illustrates the integration between the processing pipeline and the self-validation cycle, highlighting how the system approaches a paradigm of self-improving AI workflow for cybersecurity governance.

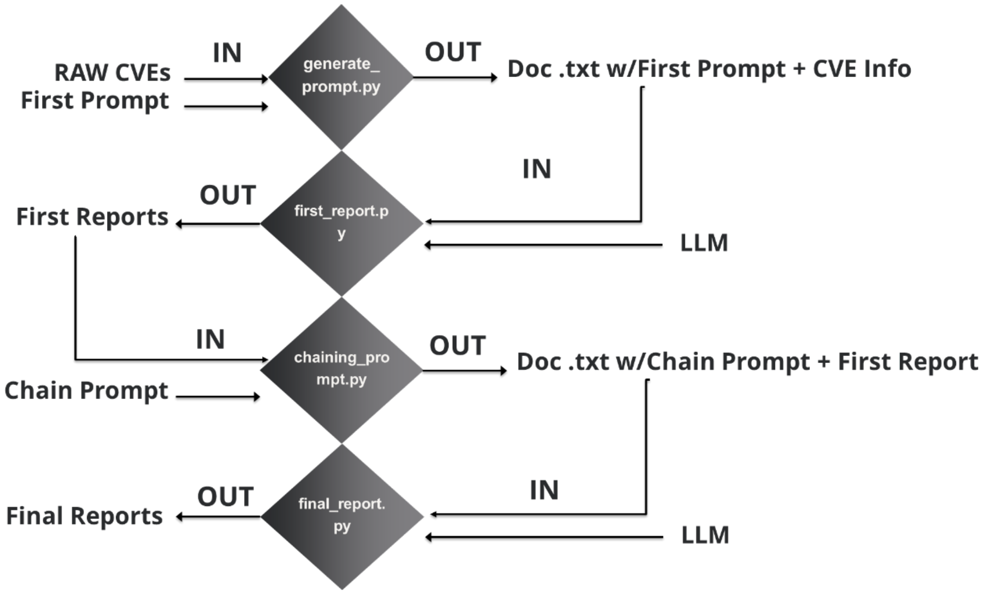


Figure 2. Processing Flow Diagram

3 Experimental campaign

The implementation of the case study aimed to verify the technical feasibility and operational performance of the proposed architecture, evaluating its ability to automate the security reporting lifecycle in a realistic context.

The test environment was configured on Apple Silicon platforms, specifically on a MacBook Pro M3 Max (14-core CPU, 30-core GPU, 32 GB RAM), to ensure efficient and stable local inference. The execution of the models was managed through the llama.cpp framework, optimized for the ARM64 architecture, using quantized models in GGUF format.

This reflects a growing trend in the adoption of on-premise AI solutions, which offer greater data security, cost reduction, and infrastructure autonomy compared to traditional cloud services.

A key aspect of the experiment was the design of the prompts, conceived to replicate the operational dynamics of a Security Operation Center (SOC).

The junior analyst's prompt presented detailed and rigid instructions, aimed at avoiding hallucinations and ensuring the presence of all the required sections (Summary, Technical Details, Risk Assessment, Suggested Mitigations, Exploitation Scenario).

The senior auditor's prompt, on the other hand, simulated a quality assurance process, aimed at improving clarity, professional tone and technical precision. This self-revision step forms the core of the prompt chaining mechanism, which allows the system to autonomously refine its textual production.

A hybrid approach was adopted for the evaluation of the results, combining quantitative and qualitative methods:

- Automatic evaluation, based on objective metrics (success rate, average generation time, completeness and presence of additional sections).
- Qualitative semi-automated assessment, conducted with GPT-4o as a proxy reviewer, which assigned scores from 1 to 5 for the criteria of completeness, clarity, structural adherence, and technical precision.

3.1 Experimental results

In the first round of generation, the models showed significant differences.

DeepSeek achieved the highest quality (score 5.0/5.0), but with significantly above-average generation times.

LilyCybersecurity and Mistral stood out for their efficiency and reliability, producing comprehensive reports with very short processing times (13.45 s and 14.06 s respectively).

WhiteRabbitNeo, while showing variability in results, provided consistent, though less structurally sound, outputs.

In the second round, based on the review phase through prompt chaining, the behavior of the models changed significantly. DeepSeek has shown a serious scalability limitation, managing to complete only 10% of the tasks due to the increasing size of the prompts.

In contrast, LilyCybersecurity has proven to be the most resilient model, successfully completing 90% of requests and maintaining high structural consistency.

Mistral and WhiteRabbitNeo benefited greatly from the iterative process, significantly improving the clarity and accuracy of the final reports.

Table 1 – Quantitative results of the models

Model	Round	Success Rate (%)	Average Time (sec)	Completeness (%)
DeepSeek	1°	100	22.80	98
Mistral	1°	95	14.06	94
WhiteRabbitNeo	1°	87	15.20	90
LilyCybersecurity	1°	98	13.45	96
DeepSeek	2°	10	—	45
Mistral	2°	85	15.75	97
WhiteRabbitNeo	2°	88	16.10	96
LilyCybersecurity	2°	90	13.90	98

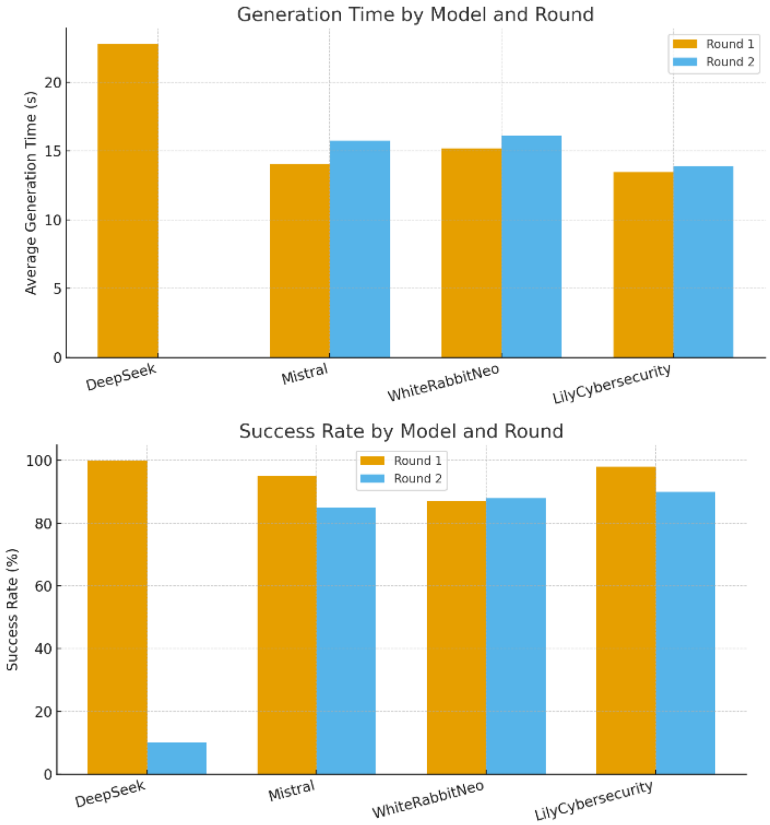


Figure 3 : Comparison between the generation times and the success rate of the models in the two rounds

Table 2. Qualitative Assessment (scores 1–5, GPT-4o)

Model	Round	Completeness	Clarity	Structural adhesion	Technical precision
DeepSeek	1°	5.0	4.8	4.9	5.0
Mistral	1°	4.3	4.0	4.1	4.0
WhiteRabbitNeo	1°	3.9	3.7	3.8	3.8
LilyCybersecurity	1°	4.8	4.5	4.6	4.7
DeepSeek	2°	2.0	2.5	2.2	2.3
Mistral	2°	4.7	4.6	4.8	4.7
WhiteRabbitNeo	2°	4.5	4.4	4.5	4.6
LilyCybersecurity	2°	4.9	4.8	4.9	4.9

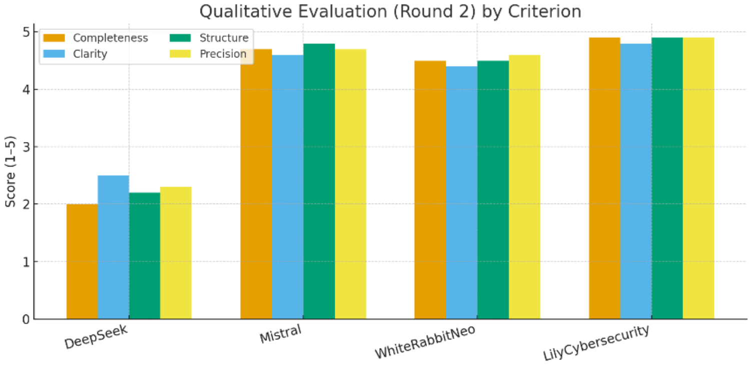


Figure 4.Average model qualitative scores for each criterion

3.2 Discussion and implications

The overall analysis shows how prompt chaining has contributed substantially to improving the clarity, structure, and technical accuracy of the generated reports.

The comparison of models also revealed some strategic evidence relevant to the governance of E-Business Information Systems (EBIS):

Trade-off between quality and scalability. The size of the model does not guarantee superior performance. DeepSeek, while excellent in single quality, does not scale in iterative pipelines; medium-sized models such as LilyCybersecurity are more balanced.

Adaptive architectures. The use of prompt chaining has proven to be essential for increasing reliability. This indicates that future EBIS architectures will need to incorporate self-review and continuous improvement mechanisms.

Compliance automation. The proposed system can automate the production of regulatory documentation (e.g. GDPR, NIS2), improving accuracy and traceability, and thus strengthening digital trust within company infrastructures.

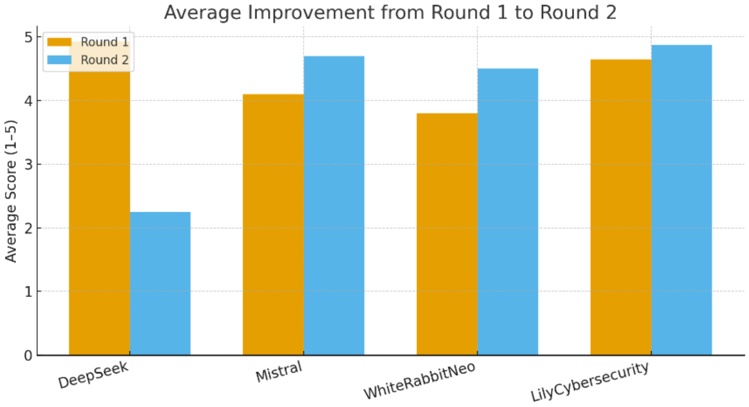


Figure 5. Comparative summary between round 1 and round 2, highlighting the average overall improvement for each model

4 Conclusions and Future Research Paradigms for the Evolution of EBIS

This study demonstrated that the automatic generation of cybersecurity reports using large language models (LLMs) represents a viable, effective and profoundly innovative strategy for the evolution of E-Business Information Systems (EBIS).

The integration of intelligent architectures based on prompt chaining has made it possible to validate a fundamental principle: efficiency and quality do not depend exclusively on the scale of the model, but rather on the design of the cognitive flow and the contextual control of the generative process.

The experimental results showed that medium-sized models, if properly guided by a structured pipeline, are able to produce consistent, technically accurate reports that comply with professional standards. In particular, LilyCybersecurity stood out as the model most balanced between speed, robustness and semantic quality, showing an operational readiness that makes it realistic to use in production scenarios.

Looking beyond the results of this study, the research paves the way towards new development paradigms for the EBIS of the future, in which artificial intelligence is not only an analytical aid, but an active component of resilience, trust and sustainability. The main perspectives are divided into four complementary directions:

1. Cyber-Physical Integration

   The first horizon concerns the extension of this approach to cyber-physical systems, such as IoT infrastructures or OT industrial environments. In such contexts, the ability to analyze vulnerabilities in real time and automatically generate contextualized reports could be a game-changer for operational resilience and real-time risk management.

2. Quantum-Safe and Quantum-Powered EBIS

   A second emerging area of research concerns the intersection between artificial intelligence and quantum computing. LLMs and hybrid AI-quantum systems will be able to support the transition to quantum-safe cryptographic algorithms, automatically identifying vulnerable libraries, legacy dependencies, or risky configurations. At the same time, the use of quantum accelerators will enhance the scalability of analysis and classification processes, paving the way for quantum-powered EBIS, capable of learning and reacting with unprecedented speed.

3. Sustainability and Green E-Business Systems

   In an era in which the environmental impact of digital technologies can no longer be overlooked, it will be crucial to investigate the energy sustainability of AI architectures applied to EBIS. The adoption of lighter models, the optimization of local inference and the integration with energy-efficient infrastructures can give rise to "green" solutions, in which IT security is combined with environmental responsibility.

4. Reinforcement Learning for Adaptive Trust

   Finally, a particularly promising prospect is that of reinforcement learning applied to adaptive trust. By training LLMs to continuously improve based on human feedback – for example, corrections or evaluations provided by analysts – it will be possible to build continuous improvement loops, in which the system learns to recognize and correct its mistakes on its own.

   This paradigm will lead to the creation of cognitively adaptive EBIS, capable of strengthening the relationship of trust between humans and machines and ensuring reliable, transparent and verifiable performance over time.

The research shows that the synergy between LLMs, reporting automation and adaptive architectures is a concrete step towards the next generation of enterprise information systems: smarter, more sustainable and resilient. The evolution of EBIS will increasingly move from a static and descriptive paradigm to a dynamic, cognitive and predictive one, capable of proactively learning, explaining and ensuring safety.

References

[1] R. Pizzolante, A. Castiglione, B. Carpentieri, R. Contaldo, G. D'Angelo, and F. Palmieri, "A machine learning-based memory forensics methodology for TOR browser artifacts," *Concurr Comput*, vol. 33, no. 23, Dec. 2021, doi: 10.1002/cpe.5935.

[2] S. Guertin-Lahoud, C. K. Coursaris, S. Sénécal, and P.-M. Léger, "User Experience Evaluation in Shared Interactive Virtual Reality," *Cyberpsychol Behav Soc Netw*, vol. 26, no. 4, Apr. 2023, doi: 10.1089/cyber.2022.0261.

[3] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. van den Berg, I. Titov, and M. Welling, "Modeling Relational Data with Graph Convolutional Networks," 2018, pp. 593–607. doi: 10.1007/978-3-319-93417-4_38.

[4] A. Gaeta, V. Loia, A. Lorusso, F. Orciuoli, and A. Pascuzzo, "Towards a LLM-based intelligent system for detecting propaganda within textual content," *Computers and Electrical Engineering*, vol. 128, p. 110765, Dec. 2025, doi: 10.1016/j.compeleceng.2025.110765.

[5] E. Bellini, G. D'Aniello, F. Flammini, and R. Gaeta, "Situation Awareness for Cyber Resilience: A review," *International Journal of Critical Infrastructure Protection*, vol. 49, p. 100755, Jul. 2025, doi: 10.1016/j.ijcip.2025.100755.

[6] T. Baker, S. Bechhofer, A. Isaac, A. Miles, G. Schreiber, and E. Summers, "Key choices in the design of Simple Knowledge Organization System (SKOS)," *Journal of Web Semantics*, vol. 20, pp. 35–49, May 2013, doi: 10.1016/j.websem.2013.05.001.

[7] M. Casillo, L. Cecere, F. Colace, A. Lorusso, D. Santaniello, and C. Valentino, "Exhibition spaces in the metaverse: a novel design approach," in *2023 8th IEEE History of Electrotechnology Conference (HISTELCON)*, IEEE, Sep. 2023, pp. 116–119. doi: 10.1109/HISTELCON56357.2023.10365847.

[8] R. J. Raimundo and A. T. Rosário, “Cybersecurity in the Internet of Things in Industrial Management,” *Applied Sciences*, vol. 12, no. 3, p. 1598, Feb. 2022, doi: 10.3390/app12031598.

[9] A. Castiglione, J. G. Esposito, V. Loia, M. Nappi, C. Pero, and M. Polsinelli, “Integrating Post-Quantum Cryptography and Blockchain to Secure Low-Cost IoT Devices,” *IEEE Trans Industr Inform*, vol. 21, no. 2, pp. 1674–1683, Feb. 2025, doi: 10.1109/TII.2024.3485796.

[10] L. Cecere, M. Grimaldi, A. Lorusso, A. Marra, and F. Stoia, “Immersive Urban Planning: Evaluating Park Safety Perception with Digital Twins and Metaverse Simulation,” *Sustainability*, vol. 17, no. 17, p. 7608, Aug. 2025, doi: 10.3390/su17177608.

[11] L. Wu, F. Petroni, M. Josifoski, S. Riedel, and L. Zettlemoyer, “Scalable Zero-shot Entity Linking with Dense Entity Retrieval,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 6397–6407. doi: 10.18653/v1/2020.emnlp-main.519.

[12] A. Della Greca, A. Ilaria, C. Tucci, N. Frugieri, and G. Tortora, “A user study on the relationship between empathy and facial-based emotion simulation in Virtual Reality,” in *Proceedings of the 2024 International Conference on Advanced Visual Interfaces*, New York, NY, USA: ACM, Jun. 2024, pp. 1–9. doi: 10.1145/3656650.3656691.

[13] A. Gaeta, V. Loia, A. Lorusso, F. Orciuoli, and A. Pascuzzo, “Computational analysis of Information Disorder in Cognitive Warfare,” *Online Soc Netw Media*, vol. 48, p. 100322, Sep. 2025, doi: 10.1016/j.osnem.2025.100322.

[14] Y. Jiang, S. Bordia, Z. Zhong, C. Dognin, M. Singh, and M. Bansal, “HoVer: A Dataset for Many-Hop Fact Extraction And Claim Verification,” in *Findings of the Association for Computational Linguistics: EMNLP 2020*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 3441–3460. doi: 10.18653/v1/2020.findings-emnlp.309.

[15] M. Casillo, F. Colace, A. Lorusso, D. Santaniello, and C. Valentino, “Integrating Physical and Virtual Experiences in Cultural Tourism: An Adaptive Multimodal Recommender System,” *IEEE Access*, vol. 13, pp. 28353–28368, 2025, doi: 10.1109/ACCESS.2025.3539205.

[16] R. Ghosh, O. Farri, H.-M. von Stockhausen, M. Schmitt, and G. M. Vasile, “CVE-LLM : Automatic vulnerability evaluation in medical device industry using large language models,” Jul. 2024.

[17] M. Fakih, R. Dharmaji, H. Bouzidi, G. Q. Araya, O. Ogundare, and M. A. Al Faruque, “LLM4CVE: Enabling Iterative Automated Vulnerability Repair with Large Language Models,” Jan. 2025.

[18] A. Lorusso and D. Guida, “IoT System for Structural Monitoring,” 2022, pp. 599–606. doi: 10.1007/978-3-031-05230-9_72.

[19] A. Gaeta, V. Loia, and F. Orciuoli, “An explainable prediction method based on Fuzzy Rough Sets, TOPSIS and hexagons of opposition: Applications to the analysis of Information Disorder,” *Inf Sci (N Y)*, vol. 659, p. 120050, Feb. 2024, doi: 10.1016/j.ins.2023.120050.

[20] U. Kulsum, H. Zhu, B. Xu, and M. d’Amorim, “A Case Study of LLM for Automated Vulnerability Repair: Assessing Impact of Reasoning and Patch Validation Feedback,” in *Proceedings of the 1st ACM International Conference on AI-Powered Software*, New York, NY, USA: ACM, Jul. 2024, pp. 103–111. doi: 10.1145/3664646.3664770.

[21] Z. Sheng, Z. Chen, S. Gu, H. Huang, G. Gu, and J. Huang, “LLMs in Software Security: A Survey of Vulnerability Detection Techniques and Insights,” Feb. 2025.

[22] S. Tian *et al.*, “Exploring the Role of Large Language Models in Cybersecurity: A Systematic Survey,” Apr. 2025.

[23] M. Casillo, L. Cecere, S. P. Dembele, A. Lorusso, D. Santaniello, and C. Valentino, “The Metaverse and Revolutionary Perspectives for the Smart Cities of the Future,” 2024, pp. 215–225. doi: 10.1007/978-981-97-3305-7_17.

[24] M. Casillo, F. Colace, A. Lorusso, D. Santaniello, and C. Valentino, “A multilevel graph approach for IoT-based complex scenario management through situation awareness and

semantic approaches," *J Reliab Intell Environ*, vol. 10, no. 4, pp. 395–411, Dec. 2024, doi: 10.1007/s40860-024-00224-0.

[25] A. A. Cantone, M. Ercolino, M. Romano, and G. Vitiello, "Designing Virtual Interactive Objects to Enhance Visitors' Experience in Cultural Exhibits," in *Proceedings of the 2nd International Conference of the ACM Greek SIGCHI Chapter*, New York, NY, USA: ACM, Sep. 2023, pp. 1–5. doi: 10.1145/3609987.3610110.