Enhancing Cybersecurity in Water Plant Infrastructure with SecureAI

Ilkka Tikanmäki[1,2]1 Diana Marinca[1], Muhibba Saani Mohammed[1] and Rakyan Kadar Slamet[1]

[1]Laurea University of Applied Sciences, Espoo, Finland
[2]National Defence University, Helsinki, Finland

Ilkka.tikanmaki@laurea.fi

Abstract: Water plant industries are a critical infrastructure that relies on legacy Supervisory Control and Data Acquisition (SCADA) systems, which were not designed to address modern cyber threats. Attackers utilise these vulnerabilities to create significant risks to the industries. As an example, recent incidents such as the attack on the Demin water plant and the Oldsmar water facility, where attackers gained unauthorised remote access to these water plants' systems. This emphasises the urgency of strengthening cybersecurity in this sector. This study investigates SecureAI, an AI-driven cybersecurity tool developed through the Dynamo project. SecureAI provides real-time anomaly detection, recommends isolating protocols to contain threats early, and generates post-incident training materials to improve operator readiness. To ensure that SecureAI implemented into critical infrastructure cannot become autonomous, the EU AI Act requires mandatory human oversight for all high-risk systems. The study includes an evaluation of SecureAI's strengths, weaknesses, and ethical safeguards that align with the EU AI Act, the NIS2 Directive, and the NIST Cybersecurity Framework. Mock-up data on attack scenarios, best practices for the deployment of SecureAI, and an incident response script designed for operator use based on SecureAI alerts. This study bridges the gap between technical detection and regulatory compliance and extends the body of knowledge on AI-enabled cybersecurity measures in water plants. Moreover, SecureAI offers a scalable, operator-centric solution that strengthens resilience while ensuring transparency, compliance and human accountability.

Keywords: SCADA, Water plant infrastructure, Human oversight, Compliance artifacts, Dynamo project

1. Introduction

Water treatment plants rely heavily on SCADA systems to ensure the safe and continuous delivery of clean water. Many of these systems, however, are based on legacy technologies that were not designed to withstand modern cyber threats. The increasing use of remote access for monitoring and maintenance has further expanded their attack surface, making water infrastructure an attractive target for threat actors (Swathika et al., 2024).

Recent incidents have demonstrated that even brief unauthorised access to control systems can pose serious risks to public health and safety. These events highlight the need for improved monitoring and stronger support for human operators, who often must find mitigation strategies during such periods.

This work-in-progress paper presents SecureAI, an operator-centred cybersecurity tool that combines real-time anomaly detection with guided response and training support. By aligning technical detection capabilities with regulatory frameworks such as the EU AI Act, NIST and NIS2, this work explores how AI-based tools can enhance the security and operational resilience of water treatment infrastructure.

This study enhances the academic literature on critical infrastructure cybersecurity. This analysis provides a systematic approach to ensuring that AI-based detection and actionable intelligence comply with the EU AI Act, the NIS2 Directive, and the NIST CSF when operating in a water SCADA/OT environment.

The rest of the paper is organised into five chapters. Chapter 2 presents a literature review, and Chapter 3 offers the method. Chapter 4 outlines the results, Chapter 5 includes a discussion, and finally, Chapter 6 offers the conclusions of the study.

2. Literature

Water treatment plants are critical infrastructure which rely on Supervisory Control and Data Acquisition (SCADA) systems for monitoring and control. Many of these installations operate on outdated or legacy operating systems, which makes them highly vulnerable to cyberattacks (Alanazi et al., 2023; Shewale, 2025) . The reliance on remote access pathways further compounds these risks, as attackers can exploit unsecured connections to bypass traditional safeguards and gain unauthorised control of plant operations.

[1]https://orcid.org/0000- 0001-8950-5221

Recent incidents highlight the urgency of addressing these vulnerabilities. In the Demin Water Plant attack, attackers exploited a remote maintenance support link to gain unauthorised access, demonstrating how weak remote access pathways can undermine security controls. The Oldsmar Water Plant attack revealed that attackers remotely accessed SCADA systems and attempted to increase sodium hydroxide levels to 100 times the normal amount. Fortunately, an alert operator noticed and reversed the change before any harm occurred (America's Cyber Defence Agency, 2021).

Both cases illustrate the risks posed by legacy SCADA systems when remote access is not properly monitored. Existing literature highlights the importance of resilience in critical infrastructure but often focuses on technical performance rather than operator usability or compliance integration. This project builds on those findings by proposing SecureAI, a tool developed by the DYNAMO project to help operators detect real-time anomalies, provide mitigation strategies and post-incident training response (DYNAMO project, 2024). This research aims to bridge the gap between technical detection and legal compliance in water plant infrastructure by aligning SecureAI with regulatory compliance, such as the EU AI Act (European Parliament, 2025), NIS2 Directive (European Parliament and Council, 2022) and NIST Cybersecurity Framework (NIST, 2024),

3. Method

The research methodology combines desktop research, technical literature review and regulatory analysis to identify gaps between technical detection and legal compliance. Building on these findings, the study proposes a mapping framework that translates SecureAI outputs, such as alerts, logs, and incident timelines, into an audit draft compliance artifact aligned with NIS2. The study used a six-phase approach over three months to investigate the integration of SecureAI into water utility SCADA environments for water plants.

The work progressed through stages:

1. Planning: The project scope, objectives, timeline, roles, and responsibilities are outlined and defined.
2. Case Study Analysis: The Demin and Oldsmar water plant attacks are reviewed and studied to understand vulnerabilities and impacts.
3. Attack Vector Investigation: Common attack vectors, particularly remote access pathways and simulated scenarios to refine SecureAI's detection parameters, are identified.
4. SecureAI Evaluation: Assess strengths, limitations and ethical safeguards of the tool.
5. Response Design: Create an incident response script reflecting oversight and responsible AI deployment
6. Documentation: Compile findings into a final report, including scripts, case study analysis and recommendations

The methodology is guided by compliance with three major frameworks: the EU AI Act, which mandates transparency, accountability, and human oversight for high-risk AI systems, the NIS2 Directive, which regulates common cybersecurity standards across EU critical entities, and the NIST Cybersecurity Framework, which provides a structured model for identifying, protecting, responding to and recovering from cyber incidents.

Structured model data based on the Demin and Oldsmar cases is used in the study, as SecureAI is not yet publicly available. The evaluation framework involves scenario-based detection testing, human-centred instruction clarity assessment, and validation of compliance mapping.

4. Results

SecureAI used the regulatory-aligned guidelines that stressed human oversight to generate containment recommendations. SecureAI is a specialised DYNAMO AI cybersecurity tool designed to identify risk, threats, and abnormalities across different data sources in water plant SCADA environments (DYNAMO project, 2024). SecureAI's architecture is closely aligned with the five pillars of the NIST Cybersecurity Framework. It can identify vulnerabilities within legacy SCADA systems and monitor remote access points, protecting assets by suggesting isolation protocols and detecting anomalies in real time through AI-driven pattern recognition. When threats are detected, SecureAI provides guided containment strategies to support a timely response, while recovery is facilitated through operator training materials and comprehensive documentation.

In addition to technical functionality, SecureAI integrates compliance mechanisms based on the EU AI Act. The system enforces strict human-in-the-loop controls, ensuring that operators retain full authority over all critical actions. It is explicitly designed not to autonomously execute irreversible decisions, thereby prioritising human

oversight and accountability. SecureAI also maintains a transparent audit trail of all actions and decisions, which is essential for post-incident review and regulatory compliance.

Deliverables from this work include a detailed evaluation report assessing SecureAI's strengths, limitations, and ethical safeguards, an incident response script tailored for operator use based on SecureAI alerts and mock-up data on attack scenarios to demonstrate optimal deployment locations. Together, these outputs demonstrate how SecureAI bridges technical detection with regulatory requirements, offering a scalable and operator-centric solution for enhancing resilience in water infrastructure.

5. Discussion

Through literature-based scenarios and compliance mapping, this work establishes SecureAI as an empirically validated product. The objective is to emphasise the social-technical requirements for operating processes and governance techniques. A primary consideration for SecureAI's success is its user interface. The tool must be intuitive and user-friendly to ensure operators can act quickly and correctly during critical incidents. Ease of use is not only a practical requirement but also an ethical one, since credibility and accountability depend on operators' ability to interact with the system effectively. The ethical implications of AI-driven decisions are significant. SecureAI's human-centric design means responsibility for authorising critical actions lies with individuals, yet audits following major incidents may target company protocols and the AI system itself. This reinforces the importance of maintaining a transparent audit trail and compliance framework.

It must be noted that SecureAI is not publicly available. Accordingly, the current analysis is based on theoretical data derived from existing information. The absence of real-world deployment data limits the ability to evaluate SecureAI's benefits compared to other AI-based security tools. Furthermore, technical details regarding installation in test environments are omitted due to a lack of direct access, requiring assumptions about deployment processes.

Despite these limitations, SecureAI demonstrates potential as an advanced Security Information and Event Management (SIEM) system with proprietary algorithms. Its strength lies in leveraging AI for pattern-based detection, enabling it to flag anomalies more quickly and reliably than human operators. While not intended to function as a traditional antivirus program, SecureAI's ability to identify subtle, emerging threats through complex pattern recognition makes it a valuable tool for enhancing resilience in water plant infrastructure.

6. Conclusions

This study addresses the urgent need to protect vulnerable SCADA systems within infrastructure. Water treatment plants are among the most essential services in modern society, yet their reliance on legacy SCADA platforms exposes them to significant cyberattack risks. Incidents such as those at Demin and Oldsmar have demonstrated how easily attackers can exploit unsecured remote access pathways to manipulate critical processes. Against this backdrop, SecureAI emerges as a timely and necessary solution, designed to strengthen resilience, reduce response times and empower operators to manage threats with confidence.

This project's success lies in its deliberate integration of technical innovation with regulatory and ethical safeguards. SecureAI is not simply a detection tool; it is a fully integrated AI-driven system that automates anomaly identification, guides containment strategies, and supports recovery through operator training and documentation. By aligning its design with the EU AI Act, the NIS2 Directive, and the NIST Cybersecurity Framework, SecureAI ensures that its deployment is not only technically effective but also legally compliant and ethically sound. This dual emphasis on innovation and compliance reflects the growing need for cybersecurity solutions to bridge the gap between technical detection and governance requirements.

Finally, this project demonstrates that SecureAI offers a scalable, operator-centric solution that safeguards water infrastructure against evolving cyber risks. By combining advanced AI capabilities with strict regulatory alignment and ethical responsibility, SecureAI helps ensure the operational continuity of essential services. The importance of these results lies not only in the technical achievements but also in the broader impact; a model for how AI can be responsibly integrated into critical infrastructure to enhance resilience, transparency, and human accountability.

Future research should focus on building a functional prototype of SecureAI and testing its detection accuracy and false-positive rates in a simulated water-utility environment. Longitudinal investigations are required to examine the impact of AI-supported case processing on the results of regulatory inspections under NIS2 and the EU AI Act.

Ethics Declaration: No human participants or personally identifiable information were involved. All data sources were publicly available.

AI Declaration: Keenious was used for literature exploration, and ChatGPT for drafting and refinement. Human authors verified all content.

References

Alanazi, M., Mahmood, A., & Chowdhury, M. J. M. (2023). SCADA vulnerabilities and attacks: A review of the state-of-the-art and open issues. *Computers & Security*, *125*, 103028. https://doi.org/10.1016/j.cose.2022.103028

America's Cyber Defence Agency. (2021). *Compromise of U.S. Water Treatment Facility* (Nos. AA21-042A; p. 4). America's Cyber Defence Agency. https://www.cisa.gov/sites/default/files/2023-04/AA21-042A_Joint_Cybersecurity_Advisory_Cyber_Actors_Compromise_U.S._Water_Treatment_Facility.pdf

DYNAMO project. (2024). *D4.1 – Initial prototypes of the cyber-threat intelligence gathering, extraction, sharing components and AI-based solutions. [*. https://horizon-dynamo.eu/wp-content/uploads/2025/02/DYNAMO-RPT-D41-V2-0.pdf

European Parliament. (2025, February 19). *EU AI Act: First regulation on artificial intelligence*. European Parliament Directorate General for Communication. https://www.europarl.europa.eu/pdfs/news/expert/2023/6/story/20230601STO93804/20230601STO93804_en.pdf

European Parliament and Council. (2022). *Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS2 Directive)*. Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02022L2555-20221227

NIST. (2024). *The NIST Cybersecurity Framework (CSF) 2.0* (Framework No. NIST CSWP 29; p. 32). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.CSWP.29

Shewale, V. (2025). Securing Legacy SCADA Systems: Practical strategies for the oil and gas industry. *World Journal of Advanced Research and Reviews*, *26*(2), 341–346. https://doi.org/10.30574/wjarr

Swathika, G., Karthikeyan, A., Karthikeyan, K., Sanjeevikumar, P., Karapparambil Thomas, S., & Babu, A. (2024). Critical review Of SCADA And PLC in smart buildings and energy sector. *Energy Reports*, *12*, 1518–1530. https://doi.org/10.1016/j.egyr.2024.07.041