DOI: 10.38146/bsz-ajia.2026.v74.i10.pp2809-2827

AI-based cybersecurity anomaly detection for OT systems in the EU Resilience Framework, an illustrative simulation-based methodological study[1]

Péter Hunorfi
PhD candidate
Óbuda University,
Doctoral School of Safety and Security Sciences
hunorfi.peter@phd.uni-obuda.hu

Abstract

Aim: To examine the possible role and constraints of Artificial Intelligence in anomaly detection for Operational Technology considering the EU critical entity resilience framework and concentrating on the European Commission's 2026 guidance.
Methodology: A focused reading of the Critical Entities Resilience Directive, CER guidance, NIS2 and the EU's AI Act is combined with an illustrative synthetic experiment. Three established detectors and a fixed-weight combination are evaluated on 7,481 aligned test timestamps.
Findings: The combined detector achieved an F1 score of 0.700 but missed many attack-labelled observations and incorrectly flagged some normal observations as anomalous. These outcomes illustrate detection limits and do not demonstrate improved service continuity, effective oversight by qualified personnel or regulatory compliance.
Value: This study links an accessible technical example with practical questions about integrated resilience measures, alert handling, accountability of re-

[1] The manuscript was submitted in Hungarian. Received: 10. 08. 2026. Revised: 05. 09. 2026. Accepted: 10. 10. 2026.

sponsible personnel and documentation. Its contribution is a bounded applied interpretation, not a new detection algorithm.

Keywords: operational technology, anomaly detection, cybersecurity, critical entities resilience, human oversight

Introduction

Digitalisation has brought operational technology (OT) into routine interaction with enterprise networks, cloud services, remote-maintenance platforms and the Industrial Internet of Things (IIoT). This incorporation makes industrial processes easier to observe and supervise, but it creates additional routes from a compromised IT environment to physical equipment. The requirement for OT security is to maintain safety, the integrity of industrial processes and uninterrupted delivery of essential services, rather than concentrate on information assets alone (Hunorfi & Farkas, 2025). Signature-based intrusion detection can be useful when a known tool, indicator or protocol violation is present. It is not that effective when the attack is unfamiliar, when an otherwise legitimate administrative command is issued with malicious purpose or when process values change slowly. In such cases, the first visible sign may be an unlikely relationship among sensors, actuators, network flows and control commands.

AI-based anomaly detection takes a different approach. It establishes a reference from observed data and gives a score to each record or short sequence based on how much it differs from the expected behaviour. Unsupervised methods identify unusual observations without labelled attacks, while reconstruction-based neural models learn relationships among several variables and evaluate whether those relationships remain reasonable over time. This is potentially useful in OT because labelled attack data are scarce and normal operation often follows recurring process cycles. A high anomaly score is just an indicator that further investigation is required. It does not mean that a cyberattack has occurred, but it can help to identify the cause of the anomaly or determine a safe response. Maintenance, changes in operating conditions or equipment faults can also produce unusual values (Ruff et al., 2021).

The publication of the European Commission's guidance under Article 13(5) of the Critical Entities Resilience (CER) Directive in July 2026 provides the main reason for revisiting this small experiment. This guidance places individual measures within an integrated resilience arrangement and gives specific attention to manual control in OT. The CER Directive, the NIS2 Directive and

the AI Act provide related but distinct legal contexts. They must not be collapsed into one substitutable set of requirements (European Commission, 2026; European Parliament & Council of the European Union, 2024, 2026). This paper is therefore presented as an illustrative simulation-based methodological study. It combines a focused review of selected EU provisions with an existing synthetic OT experiment using Isolation Forest (IF), a dense autoencoder, a long short-term memory (LSTM) autoencoder and a fixed-weight combination of these detectors. The purpose is to show what the recorded numerical results can and cannot contribute to a discussion of human responsibility, monitoring, alert management and documentation. The experiment does not measure resilience, legal compliance or service continuity, and it does not offer a new detection algorithm or a general ranking of model families. The selected resilience and AI-governance considerations are connected to transparent observations and limitations from a small technical example. Earlier studies already compare anomaly detectors and discuss evaluation methods. The study's novelty lies in this focused interpretation in the light of the Commission's 2026 guidance, not in the synthetic data alone or in the underlying algorithms.

The EU-related focus of the study

Resilience and the scope of the CER Directive

The CER Directive concerns the ability of identified critical entities to maintain essential services. Article 13 requires Member States to ensure that critical entities take appropriate and proportionate technical, security and organisational measures based on risk assessment and document those measures in a resilience plan or equivalent documentation. A detector can be relevant to parts of this wider arrangement, but one monitoring tool cannot represent the entity's complete resilience capability (European Parliament & Council of the European Union, 2022b, Article 13). The boundary with cybersecurity legislation matters. Article 1(2) of the CER Directive excludes matters covered by Directive (EU) 2022/2555, the NIS2 Directive, without prejudice to Article 8 of CER, and requires coordinated implementation in view of the relationship between physical security and cybersecurity. This paper therefore discusses the possible contribution of OT monitoring to cybersecurity resilience. It does not present CER as the sole source of cybersecurity obligations or indicate that AI monitoring is generally mandatory (European Parliament & Council of the European Union, 2022a, 2022b).

Human involvement as addressed in the EU guidance

The Commission's 2026 communication was issued under Article 13(5) of the CER Directive. It does not create a new legal requirement. However, it gives practical guidance on measures that critical entities may take. Point 18 recommends integrating resilience measures rather than implementing them as isolated protections. The common thread of the present study is that an anomaly detector can only be considered as one factor within wider technical, organisational and human arrangements (European Commission, 2026, points 4 and 18). The guidance supports a measurable approach also. Point 20 discusses systematic incident reporting and records. Point 21 refers to appropriate training and oversight, clear accountability, audits and process documentation. These considerations do not validate a detector, but they make the attributes of an alert relevant to organisational practice (European Commission, 2026, points 20–21). For OT, point 23 specifically recommends manual overrides and human-in-the-loop provisions that specify the criteria and requirements for intervention and manual control. The experiment in this article is advisory only as it produces scores and binary flags, without changing a valve, pump or other physical process. The guidance therefore provides a reason to examine how a flag could enter a human decision process, but the simulation does not test the effectiveness of that process (European Commission, 2026, point 23). Point 29 discusses functional testing and the analysis of legitimate detections, system faults and false alarms in the context of physical protection and perimeter monitoring (European Commission, 2026, point 29). The understanding adopted here is that monitoring needs a defined organisational purpose. Someone must assess an indication and decide whether further inspection or action is required in accordance with agreed responsibilities. This does not mean every anomaly must lead to intervention, but the route from a technical indication to a decision should be defined and documented.

The relevance of the AI Act of the EU

The AI Act requires attention to the system's intended purpose. Annex III, point 2, covers AI systems intended as safety components in the management and operation of specified critical infrastructure. Recital 55 states that components intended solely for cybersecurity purposes should not qualify as safety components. An anomaly detector is therefore not automatically a high-risk AI system simply because it is used in an OT environment. Its intended purpose, safety-related function and the applicable classification rules under Article 6 must be assessed for the actual deployment, taking account of the amendments in-

troduced by Regulation (EU) 2026/1744 (European Parliament & Council of the European Union, 2024, 2026). For systems to which the high-risk requirements apply, Articles 11–15 address technical documentation, record-keeping, transparency, human oversight, accuracy, robustness and cybersecurity. These requirements must not be attributed without distinction to every cybersecurity tool. In this study, the models produce scores and binary indications but do not issue commands to equipment. Advisory outputs and the absence of direct actuation do not, by themselves, exclude high-risk classification. The experiment's function does not determine the legal classification of a future product or installation. The timing of application must also be distinguished from the content of the requirements. Under the amended Article 113, the requirements discussed here apply from 2 December 2027 for systems classified under Article 6(2) and Annex III, and from 2 August 2028 for systems classified under Article 6(1) and Annex I (European Parliament & Council of the European Union, 2024, 2026).

Background and related work

OT security constraints

OT differs from conventional information technology because a control command can change the state of machinery, production equipment or another physical process. Monitoring arrangements must accommodate deterministic traffic, legacy protocols, narrow maintenance windows, long equipment lifecycles and strict availability requirements. NIST SP 800-82 Revision 3 accordingly states that OT safeguards must consider processing requirements, dependable operation and process safety in addition to conventional cybersecurity objectives (Stouffer et al., 2023). False alarms carry a substantial operational cost in OT environments. Too many alerts may lead to alarm fatigue, while an unsuitable automated response may interrupt production or destabilise a safety-critical process. An anomaly score should therefore support an analyst or operator decision and should not, by itself, trigger an unconditional process shutdown. This general distinction is preserved as a basic premise of the study.

Isolation Forest

IF repeatedly partitions the data into progressively smaller subsets and measures how many divisions are needed to isolate each observation from the

others. A record isolated after relatively few random splits receives a higher anomaly score. The technique is computationally economical and does not need attack labels, which makes it well suited as a reference method for OT monitoring (Liu et al., 2008). Its point-oriented operation can, however, miss cases in which individual measurements remain plausible but their temporal order or cross-signal relationships are abnormal. In this experiment, temporal summaries are included among its inputs. The reported results therefore belong to that implemented configuration rather than to the method in general.

Autoencoders and Long Short-Term Memory (LSTM) networks

An autoencoder uses an encoder to reduce the input to a lower-dimensional representation and a decoder to rebuild the original values. After the autoencoder has been trained on normal-operation records, attacks and faults may produce larger reconstruction errors because the observed relationships do not match the representation learned by the network. The method reduces the complexity of nonlinear data and identifies departures from a learned reference (Hinton & Salakhutdinov, 2006). LSTM networks use a gated memory mechanism that helps them learn patterns and associations across multiple time steps. Accordingly, an LSTM autoencoder reconstructs an entire sequence instead of an individual observation. This is useful when attacks are reflected in sustained behaviour, the order of events, or changes in process-variable rates, rather than in a single unusually high or low measurement (Hochreiter & Schmidhuber, 1997). It also creates evaluation choices at the boundaries of labelled intervals, which are made explicit below.

Positioning of the case study

Research using physical water-treatment testbeds provides a different level of evidence from a signal generator. SWaT supports work with recorded sensor and actuator data under controlled conditions (Goh et al., 2017; Mathur & Tippenhauer, 2016). Canonico et al. (2025) compare several neural anomaly detectors using network and physical data and show that the usefulness of a data source hinges on the attack being considered. Their work highlights the need to consider the available information and evaluation context instead of presuming that a single detector is appropriate for every threat. Simulation-based comparisons also exist. Sobonski and Roedig (2024) use an autoclave example to compare anomaly detectors, consider changing background operating

circumstances and assess detection at different levels, including delay. Sørbø and Ruocco (2024) explain why evaluation measures may be interpreted differently in time-series anomaly detection. Ha et al. (2022) discuss explainability in ICS anomaly detection. These studies motivate careful interpretation and operational context, and they also show why familiar models on synthetic data are not, by themselves, an algorithmic novelty.

Research Questions

The present study asks three limited questions.

RQ1: Which selected provisions of the EU resilience framework are relevant when considering an advisory AI-based anomaly detector for OT cybersecurity monitoring?

RQ2: What do the existing synthetic results illustrate about the detection of deviations associated with simulated cyberattack scenarios, missed attack-labelled timestamps and false positives during normal operation?

RQ3: Which methodological, human and organisational limitations must be recognised before the detector's output could support the assessment of potential OT cybersecurity incidents within a wider resilience process?

The document analysis addresses RQ1 by examining the selected provisions relevant to OT cybersecurity monitoring. The recorded synthetic experiment addresses RQ2 by evaluating detections, missed attack-labelled timestamps and false positives. Their combined interpretation addresses RQ3 by identifying limitations that must be considered before anomaly indications could support cybersecurity incident assessment and resilience-related decisions.

Study design

The first part is a focused document analysis of resilience measures, monitoring, human involvement and intended use in the CER Directive, the 2026 guidance and the AI Act, including the amendments introduced by Regulation (EU) 2026/1744. NIS2 clarifies the CER boundary. Selection follows the study's purpose; the analysis is not a systematic review of all EU and national requirements. The legal analysis was finalised on 30 August 2026 and reflects the legislation and guidance available as of that date. The second part presents the original synthetic experiment. Numerical findings are distinguished from practical interpretation. No new training, delay analysis, event aggregation or uncertainty estimation is introduced. Relating a result to a governance question does not demonstrate compliance.

Synthetic data and chronological separation

The generator creates 18,000 one-second records, representing five hours of synthetic time. Eight variables represent process, actuator, network and command activity (Table 1). Periodic variation, three operating modes, relationships between variables and noise create the signals. They are neither plant measurements nor captured network traffic, and the generator is not physically validated.

Table 1
Variables in the synthetic example

Variable	Meaning in the generator
pressure_bar	Process pressure
flow_lpm	Process flow rate
tank_level_pct	Tank level
temperature_c	Process temperature
valve_position_pct	Valve position
pump_speed_pct	Pump speed
network_packets_s	Network packet rate
command_rate_s	Control-command rate

Note. The names and units describe the constructed signals. They do not imply physical validation. Source: Author's original dataset and generator.

The chronological split contains 8,000 training, 2,500 validation and 7,500 test records. Training and validation are normal-only; six attack scenarios appear in the test period. The generator supplies the labels. The models receive no attack labels during fitting or threshold selection. This curated setting is called normal-only, one-class or novelty detection. Table 2 describes changes to generated signals intended to illustrate possible observable effects of cyberattacks, rather than complete attacks on a PLC or network. The attack labels reflect the scenario design; they do not mean that the detectors identify malicious intent. Sustained process manipulation starts with a step change and then varies; it is not a gradual-onset attack.

Table 2
Attack scenarios in the test period

Scenario	Interval (s)	Constructed change
Sensor bias	10,950–11,319	Pressure and tank-level offsets increase over the interval.
Actuator manipulation	11,950–12,299	Valve values change, with much smaller changes in flow and pressure.
Replay	13,200–13,549	Six process/actuator channels reuse values from seconds 1,200–1,549, with noise added.
Denial of service	14,400–14,599	Packet rate rises and command rate is reduced.

Scenario	Interval (s)	Constructed change
Sustained process manipulation	15,500–16,079	Flow, pressure, tank level and valve position are modified together.
Command injection	16,900–17,119	Command-rate and valve-position changes include bursts.

Note. Time is measured from the start of the complete series, and both listed endpoints are included. Source: Author's original generator.

Each scenario occurs once: 2,070 attack-labelled timestamps represent six intervals. Normal mode changes are included, but separate maintenance and equipment-fault scenarios are absent. The case therefore does not test whether a detector can distinguish malicious intent from all legitimate or faulty operation.

Model fitting and inputs

All variables are standardised using means and standard deviations from the training period. The same transformation is applied to validation and test records. Isolation Forest receives 24 features: the eight current values, their first differences and deviations from trailing 20-sample means. The temporal features are calculated chronologically from current and past observations across the series. The 24 features undergo a second standardisation fitted only on the training portion. The dense autoencoder receives the eight current values. The LSTM autoencoder receives the most recent 20 records as a sequence. Table 3 summarises the configurations.

Table 3
Implemented detector configurations

Detector	Input and configuration	Score
Isolation Forest	24 features; 400 trees; max_samples = 512; contamination = auto; random_state = 42.	Negative of scikit-learn score_samples; higher means more unusual.
Dense autoencoder	8–12–4–12–8 layers; ReLU in the three hidden layers; linear output. Maximum 40 epochs.	Mean squared reconstruction error for the current record.
LSTM autoencoder	20 × 8 sequence; encoder hidden size 28; latent size 12; decoder hidden size 28; linear output. Maximum 35 epochs.	Mean squared reconstruction error across the complete window.
Combined detector	Fixed score weights: 0.15 Isolation Forest, 0.40 dense autoencoder, 0.45 LSTM autoencoder.	Weighted sum of scores scaled using the validation median and 95th percentile, as described in the text.

Note. These are configurations of established methods, not new model architectures. Source: Author's original Python implementation.

The neural models minimise mean squared reconstruction error and use early stopping on normal validation data. The LSTM reconstructs a 20-step window. LSTM windows are constructed separately within each partition. Training windows use stride three, producing 2,661 windows; validation and test windows use stride one, producing 2,481 and 7,481 windows, respectively. Dense-autoencoder early stopping uses all 2,500 validation records. The implementation uses one fixed seed (42), so variation across runs is unknown. Both neural models use Adam with a learning rate of 0.001, weight decay of 0.00001 and shuffled batches of 128. Early stopping uses a patience of five epochs and a minimum validation-loss improvement of 0.00001; the best saved model state is restored. LSTM training additionally uses gradient clipping with a maximum norm of 1.0. The inputs differ: Isolation Forest has engineered temporal summaries, the dense model sees one record, and the LSTM sees a sequence. Results describe the implemented configurations. They do not isolate the effect of architecture or temporal modelling.

Thresholds and evaluation

Each threshold is the 97.5th percentile of 2,481 aligned normal validation scores, including for the combination. This places it near the upper end of normal validation scores. It is not selected using test attack labels and does not guarantee the same exceedance rate on new data. Before combination, each detector's scores are centred by subtracting the median of its 2,481 aligned normal validation scores. The centred scores are divided by the difference between the 95th percentile and the median of those validation scores, with 0.000000001 added to the denominator. These validation-derived constants are also applied to test scores, without clipping. The resulting score series are combined using the fixed weights in Table 3. A timestamp is flagged only when its score exceeds the threshold. Alternative weights and thresholds were not tested, and optimality is not claimed. The normal validation period supports both early stopping and threshold calibration, which is a limitation. The LSTM score covers 20 records but receives its final timestamp's label. All outputs are aligned to complete windows, excluding the first 19 test records. Evaluation therefore uses 7,481 timestamps: 5,411 normal and 2,070 attack-labelled. A decision concerns one timestamp, not a complete attack or operator-facing alarm. Windows at the start or end of an attack may contain both normal and attacked records. These boundary windows remain in the evaluation. No tolerance, relabelling or alarm-merging rule is applied.

Measures and the recorded results

Precision is the fraction of flagged timestamps with an attack label. Recall is the fraction of attack-labelled timestamps that are flagged. F1 is the harmonic mean of precision and recall. The false-positive rate is the part of normal timestamps incorrectly flagged. Confusion counts allow these four measures to be checked directly. Recall is also reported within each attack interval. It determines the portion flagged, rather than crediting a whole attack after one detection. Results are descriptive. Confidence intervals and significance claims are omitted because timestamps are related and independent repetitions are absent (Sørbø & Ruocco, 2024).

Results of the illustrative experiment

Overall observations

Table 4 shows the validation-calibrated results. The combined detector has the highest observed F1, 0.700. The dense autoencoder has marginally higher precision, while Isolation Forest has the lowest recall and false-positive rate. These are trade-offs within this case.

Table 4
Timestamp-level performance in the recorded test run

Detector	Precision	Recall	F1	False-positive rate
Isolation Forest	0.550	0.058	0.105	0.018
Dense autoencoder	0.895	0.475	0.621	0.021
LSTM autoencoder	0.856	0.454	0.593	0.029
Combined detector	0.893	0.576	0.700	0.026

Note. Values are reported to three decimal places. All results refer to the same 7,481 test timestamps.
Source: Author's original experiment.

Table 5 shows that the combination flags 1,192 of 2,070 attack-labelled timestamps and misses 878. It incorrectly flags 143 of 5,411 normal timestamps. These are counts of seconds, not incidents, separate alarms or operator decisions.

Table 5
Correct and incorrect timestamp decisions

Detector	Normal, not flagged	Normal, flagged	Attack, missed	Attack, flagged
Isolation Forest	5,313	98	1,950	120
Dense autoencoder	5,295	116	1,086	984
LSTM autoencoder	5,253	158	1,131	939
Combined detector	5,268	143	878	1,192

Source: Author's original experiment.

About 28% of evaluated timestamps carry attack labels. Precision depends partly on how common attacks are in the data; thus, the observed value is not an estimate for an installation with rare attacks (Axelsson, 2000). Assuming that recall and the false-positive rate remain unchanged, the observed rates imply a precision of approximately 18% if 1% of timestamps are attack-labelled, and about 2.1% if 0.1% are attack-labelled. This hypothetical calculation illustrates the base-rate effect; it is not a prediction of performance in an operational installation (Axelsson, 2000).

Differences between scenarios

Table 6 shows recall within each scenario. The dense autoencoder model flags most sensor-bias observations. The LSTM and combination flag all denial-of-service and command-injection timestamps. The findings do not show that AI is needed to detect these explicit signal changes.

Table 6
Recall within each constructed attack interval

Scenario	Seconds	IF	Dense AE	LSTM AE	Combined
Sensor bias	370	0.005	0.978	0.543	0.924
Actuator manipulation	350	0.003	0.309	0.420	0.429
Replay	350	0.023	0.220	0.489	0.551
Denial of service	200	0.275	1.000	1.000	1.000
Sustained process manipulation	580	0.021	0.224	0.000	0.150
Command injection	220	0.191	0.486	1.000	1.000

Note. IF = Isolation Forest; AE = autoencoder. A value of 1.000 means every timestamp in that one interval was flagged. Source: Author's original scenario results.

For the combined detector, the unweighted mean of the six scenario-level recall values is 0.676, compared with the overall timestamp-level recall of 0.576. This difference reflects the unequal durations of the constructed attack intervals and shows that the aggregate result is influenced by scenario duration. For

sustained process manipulation, the combination flags 15% of the recorded time points, while the LSTM flags none. This describes poor coverage of one generated interval at the selected thresholds, not performance against gradual or stealthy attacks in general. The overall F1 value thus coexists with strong detection of some changes and many missed observations on others. One interval per scenario cannot establish expected performance for an attack type or how quickly an operator would receive a useful alarm.

Discussion

The analysis addressing RQ1 distinguishes the resilience context, practical guidance and AI intended-use considerations relevant to advisory OT cybersecurity monitoring. The results addressing RQ2 show uneven detection of deviations associated with simulated cyberattack scenarios, alongside missed attack-labelled timestamps and false positives during normal operation. The interpretation addressing RQ3 identifies methodological, human and organisational limitations that must be considered before these indications could support the assessment of potential OT cybersecurity incidents within a wider resilience process. This connection reflects the author's interpretation, not a measured improvement in incident assessment or resilience. Table 7 separates case observations from practical questions recommended by the author. This is an analytical aid, not a compliance assessment or a validated control framework.

Table 7
Connections between the EU Context and the Illustrative Case

Document consideration	Observation in this case	Practical question
CER Article 13: resilience measures, documentation.	The detectors produce anomaly indications. Service continuity and recovery are not measured.	Which operational decision would the indication inform?
2026 guidance, point 18: integrated resilience measures.	The detector is assessed in isolation from other safeguards and response arrangements.	How would the indication connect to existing technical and organisational measures?
2026 guidance, points 20–21: records, accountability and documented processes.	The case produces timestamp flags, but no organisational incident record or review workflow.	What information, decision and disposition should be recorded, and by whom?
2026 guidance, point 23: human involvement and manual control in OT.	No operator response or manual override is implemented or tested.	Who may decide and act, and through which established procedure?

Document consideration	Observation in this case	Practical question
2026 guidance, point 29: testing and analysis of false positives in its stated physical-protection context.	Normal timestamps are also flagged. Event-level alarm burden is unknown.	Could an analogous review process distinguish useful detections, faults and false alarms?
AI Act, as amended: intended purpose, safety-component role and classification rules.	The experiment scores data and does not control equipment. This does not, by itself, determine legal classification.	What intended purpose, safety-related function and influence on decisions would the proposed deployment have under Article 6 and the relevant annexes?

Note. Source: Author's interpretation of the case alongside European Commission (2026) and European Parliament and Council of the European Union (2022b, 2024, 2026). The legal status and scope of each document remain distinct.

From an unusual value to a verified concern

The value of a flag depends on context. An operator might check current measurements, maintenance activity, authorised work and control-system logs. These uses are not tested here, and the model alone cannot distinguish failure, legitimate intervention and hostile actions. A proportionate arrangement would define who receives an indication, what background is presented and when escalation is applicable. The author interprets the guidance as a reason to document this decision route rather than treat a threshold crossing as a complete response (European Commission, 2026). The false-positive rate cannot establish acceptable workload. Consecutive flagged seconds may describe one continuing situation, while separated flags may lead to separate checks. Alarm formation and investigation effort were not measured. These would need evaluation with the intended users.

Operator accountability if monitoring is automated

The proposed role is advisory. The detector supplies information for review and has no direct path to change a valve, stop a pump or alter a safety function. This authorial design proposal reflects the case's limits and established OT safety concerns; it has not been validated in an operating plant (Stouffer et al., 2023). Human involvement requires information, authority and an understood response procedure. This study did not test how operators would interpret the alerts or respond to them. The models also have no tested explanation module. Explainability research identifies a relevant direction for further work (Ha et al., 2022). Any proposed deployment should be assessed against its intended

purpose, safety-related function and the applicable AI Act provisions. Neither use in a critical sector nor an advisory role alone settles its classification (European Parliament & Council of the European Union, 2024, 2026).

Documentation and continued review

A useful record would identify the normal reference data, detector version, threshold and intended use. Operating-mode, equipment or data-collection changes may alter what the model considers unusual. Such changes should be recorded and performance reviewed with relevant site data. One fixed series cannot show that a threshold remains suitable over time (Tabassi, 2023). The detector and its data pipeline also require protection. Manipulated inputs or contaminated reference data can undercut machine-learning systems (Vassilev et al., 2025). The manipulation resistance is not tested here, so the monitoring would need its own risk assessment.

Limitations

The case contains one synthetic series, split and fitting run. Five hours and six attack intervals do not represent variation between plants, equipment states or attackers. Signals are not physically validated, and labels are less ambiguous than in practice. The findings cannot establish production detection rates or improved service continuity. Inputs differ and weights and thresholds are fixed. Simple process-rule baselines, matched-input comparisons and sensitivity analyses are absent. The case cannot establish that neural models are necessary, the combination is optimal or a model family is generally better. Timestamp evaluation does not measure incidents, delay or operator workload. Window labels can affect boundary results. Repeated-run variability, prediction time, memory use and edge-device suitability were not assessed. Dependent seconds are not treated as independent evidence. The selective EU-level analysis does not cover all sectoral rules, national implementation or deployment obligations. Table 7 has not been validated with operators, exercises or a compliance assessment. Manual intervention is discussed as a practical consideration, not as a tested capability. Operational adoption would require further evidence, including independent cyber-physical data, benign disturbances and evaluation with users. The present case supports discussion and preparation within its stated limits.

Conclusion

The CER Directive, the 2026 guidance and the AI Act provide some perspectives on the potential role of AI-based OT monitoring. Their interpretation highlights alert verification, responsibility and documentation. It does not establish a general obligation to introduce AI or compliance by the models studied. The synthetic case shows both correctly flagged attack-labelled timestamps and detection errors. The combination's F1 of 0.700 and the uneven scenario results support a discussion of limitations, without demonstrating general superiority or operational readiness. The contribution is an illustrative interpretation connecting the selected EU context to the original experiment. It distinguishes the information a detector supplies from the human and organisational decisions that remain necessary, and from the capabilities the case has not tested.

References

Axelsson, S. (2000). The base-rate fallacy and the difficulty of intrusion detection. *ACM Transactions on Information and System Security, 3*(3), 186–205. https://doi.org/10.1145/357830.357849

Canonico, R., Esposito, G., Navarro, A., Romano, S. P., Sperlí, G., & Vignali, A. (2025). An anomaly-based approach for cyber–physical threat detection using network and sensor data. *Computer Communications, 234*, Article 108087. https://doi.org/10.1016/j.comcom.2025.108087

European Commission. (2026). Communication from the Commission: Guidelines on the application of Article 13(5) of Directive (EU) 2022/2557 on the resilience of critical entities. *Official Journal of the European Union*, C, C/2026/3712. https://eur-lex.europa.eu/eli/C/2026/3712/oj/eng

European Parliament & Council of the European Union. (2022a). Directive (EU) 2022/2555 of 14 December 2022 on measures for a high common level of cybersecurity across the Union. *Official Journal of the European Union*, L 333, 80–152. https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng

European Parliament & Council of the European Union. (2022b). Directive (EU) 2022/2557 of 14 December 2022 on the resilience of critical entities and repealing Council Directive 2008/114/EC. *Official Journal of the European Union*, L 333, 164–198. https://eur-lex.europa.eu/eli/dir/2022/2557/oj/eng

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L, 2024/1689. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

European Parliament & Council of the European Union. (2026). Regulation (EU) 2026/1744 of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI). Official Journal of the European Union, L, 2026/1744. https://eur-lex.europa.eu/eli/reg/2026/1744/oj/eng

Goh, J., Adepu, S., Junejo, K. N., & Mathur, A. (2017). A dataset to support research in the design of secure water treatment systems. In G. Havarneanu, R. Setola, H. Nassopoulos, & S. Wolthusen (Eds.), *Critical information infrastructures security* (Lecture Notes in Computer Science, Vol. 10242, pp. 88–99). Springer. https://doi.org/10.1007/978-3-319-71368-7_8

Ha, D. T., Hoang, N. X., Hoang, N. V., Du, N. H., Huong, T. T., & Tran, K. P. (2022). Explainable anomaly detection for industrial control system cybersecurity. *IFAC-PapersOnLine, 55*(10), 1183–1188. https://doi.org/10.1016/j.ifacol.2022.09.550

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science, 313*(5786), 504–507. https://doi.org/10.1126/science.1127647

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hunorfi, P. (2026). AI-based cybersecurity anomaly detection for OT systems in the EU Resilience Framework Dataset (Version 1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.23450323

Hunorfi, P., & Farkas, T. (2025). Az operatív technológia kiberbiztonsága kritikus infrastruktúrákban [Cybersecurity of operational technology in critical infrastructure]. *Belügyi Szemle, 73*(SI1), 65–81. https://doi.org/10.38146/bsz-ajia.2025.v73.i1si.pp65-81

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining* (pp. 413–422). IEEE. https://doi.org/10.1109/ICDM.2008.17

Mathur, A. P., & Tippenhauer, N. O. (2016). SWaT: A water treatment testbed for research and training on ICS security. In *2016 International Workshop on Cyber-Physical Systems for Smart Water Networks* (pp. 31–36). IEEE. https://doi.org/10.1109/CySWater.2016.7469060

Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., Dietterich, T. G., & Müller, K.-R. (2021). A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE, 109*(5), 756–795. https://doi.org/10.1109/JPROC.2021.3052449

Sobonski, P., & Roedig, U. (2024). Comparison of industrial control system anomaly detection methods. In *Proceedings of the 2024 Workshop on Re-design Industrial Control Systems with Security* (pp. 79–85). Association for Computing Machinery. https://doi.org/10.1145/3689930.3695211

Sørbø, S., & Ruocco, M. (2024). Navigating the metric maze: A taxonomy of evaluation metrics for anomaly detection in time series. *Data Mining and Knowledge Discovery, 38*, 1027–1068. https://doi.org/10.1007/s10618-023-00988-8

Stouffer, K., Pease, M., Tang, C., Zimmerman, T., Pillitteri, V., Lightman, S., Hahn, A., Saravia, S., Sherule, A., & Thompson, M. (2023). Guide to operational technology (OT) security (NIST SP 800-82 Rev. 3). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-82r3

Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., & Hamin, M. (2025). *Adversarial machine learning: A taxonomy and terminology of attacks and mitigations* (NIST AI 100-2e2025). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-2e2025

Recommended citation

Hunorfi, P. (2026). AI-based cybersecurity anomaly detection for OT systems in the EU Resilience Framework, an illustrative simulation-based methodological study. *Belügyi Szemle, 74*(10), 2809–2827. https://doi.org/10.38146/bsz-ajia.2026.v74.i10.pp2809-2827

Declarations

Conflict of interest

The author declared no conflicts of interest.

Funding

The author received no specific funding for this research.

Ethics statement

The author declares that the study was conducted in accordance with the applicable research ethics and publication ethics requirements. The study used synthetic data generated by the Python implementation described in the article.

Data and code availability

The synthetic dataset, Python implementation, recorded aggregate and scenario-level results, and accompanying README are publicly available in version 1.0 of the supplementary package deposited on Zenodo (Hunorfi, 2026): https://doi.org/10.5281/zenodo.23450323.

The dataset contains all 18,000 records, including chronological partition assignments and attack labels. The experiment was conducted using Python 3.14.7 (64-bit), PyTorch 2.13.0+cpu and scikit-learn 1.9.0. Additional dependencies included NumPy 2.5.2, pandas 3.0.5 and SciPy 1.18.1.

Use of artificial intelligence

OpenAI's ChatGPT was used to assist with translation into English, language editing, grammatical correction, manuscript restructuring and formatting, reference-format checking, and the

development of parts of the accompanying Python code. The author reviewed the suggestions, sources, code, calculations and interpretation. The author retains full responsibility for the integrity of the submitted content and for compliance with research ethics requirements, COPE guidance and the journal's policies.

Open access

This article is published under the terms of the Creative Commons Attribution–NonCommercial–NoDerivatives 4.0 International License (CC BY-NC-ND 4.0). The article may be freely shared and redistributed in any medium or format for non-commercial purposes, provided that it remains unchanged and appropriate credit is given to the original author and the source of publication, together with a link to the licence. Distribution of modified or adapted versions and commercial use are not permitted under this licence.

Corresponding author

The corresponding author of this article is Péter Hunorfi, who can be contacted at hunorfi.peter@phd.uni-obuda.hu.