DOI: https://doi.org/10.15276/hait.09.2026.46
UDC 004.85:004.62:351.86

Real-time extraction of cognitive-semantic security indicators: a natural-language-processing pipeline for anomaly-driven adaptive governance

Maksym A. Rohovets[1)]
ORCID: https://orcid.org/0000-0002-1587-9017; mrohovets@gmail.com
Victor V. Ievdokymov[2)]
ORCID: https://orcid.org/0000-0002-3577-081X; viktorievdokymov@gmail.com
[1)] Defense Intelligence Research Institute, 81, Yurii Illienko Str. Kyiv, 04050, Ukraine
[2)] Zhytomyr Polytechnic State University, 103, Chudnivska Str. Zhytomyr, 10005, Ukraine

ABSTRACT

Relevance. Adaptive governance of the security environment under hybrid aggression depends on converting large, noisy, natural-language information flows into machine-computable indicators fast enough to act before a threat materializes. Prior cognitive-semantic models parametrize the security graph manually from expert elicitation, which does not scale to real-time monitoring. **Aim.** To design and validate an end-to-end natural-language-processing pipeline that automatically extracts cognitive-semantic risk indicators from a text stream, detects semantic anomalies, and drives a downstream cognitive map to produce an integral state-security indicator in near real time. **Methods.** The pipeline comprises a sequence of stages: ingestion and Ukrainian-language preprocessing; concept mapping onto a defined set of security concepts using term frequency–inverse document frequency (TF-IDF) semantic anchors and cosine similarity; per-concept indicator scoring combining a destructive-polarity lexicon with document volume; anomaly detection using a one-sided sliding-window z-score corroborated by an Isolation Forest; and driving a weighted cognitive map whose adjacency structure is adopted from the authors' prior model, with an anomaly-triggered adaptive-governance response. The pipeline is exercised end-to-end on a structured demonstration corpus of Ukrainian open-source intelligence (OSINT)-style records organized into daily buckets and containing an injected combined-escalation event. **Results.** Concept mapping reproduced the reference labelling with a high level of agreement. Anomaly detection raised the first flag ahead of the modelled escalation peak, providing an actionable early-warning window. Driven by the extracted indicators, the cognitive map reproduced the expected divergence: without an adaptive response the integral indicator collapsed to a critically low fraction of its normative level, whereas the anomaly-triggered adaptive branch arrested the decline and recovered the system to a stable plateau. **Conclusions.** Automating indicator extraction closes the gap between raw information flows and the cognitive model, turning a static expert-parametrized construct into a monitoring instrument that can be updated continuously. The pipeline is modular, reproducible, and portable to an operational stream; its main dependencies are the quality of Ukrainian-language linguistic resources and the calibration of the anomaly thresholds.

Keywords: Natural language processing; text mining; anomaly detection; Isolation Forest; cognitive map; security environment; adaptive governance

***For citation*: Rohovets M. A., Ievdokymov V. V. "Real-time extraction of cognitive-semantic security indicators: a natural-language-processing pipeline for anomaly-driven adaptive governance". *Herald of Advanced Information Technology*. 2026; Vol.9 No.4: 752–759. DOI: https://doi.org/10.15276/hait.09.2026.46**

INTRODUCTION

The contemporary security environment of Ukraine is a highly dynamic, multidimensional system shaped by the ongoing armed aggression of the Russian Federation and the systematic use of hybrid instruments of influence [1], [2]. The convergence of military, cyber, informational, economic, and political pressure produces an unprecedented configuration of risks whose management requires processing fragmented information flows faster than adversary action unfolds [3], [4].

A cognitive-semantic representation of the security environment – a weighted directed graph over a small set of interacting concepts – has proven effective for capturing nonlinear cascades and hidden causal links that classical probabilistic models handle poorly [5], [6]. In the authors' prior work, such a model was formalized and validated on expert-elicited weights [7]. However, its principal operational limitation is that both the driving impulses and the moment of intervention are set manually. This decouples the model from the live information stream and prevents its use as a real-time monitoring instrument.

This paper addresses that limitation from an information-technology standpoint. Rather than proposing a new governance model, it contributes the computational pipeline that feeds one: an automated chain that converts a Ukrainian-language text stream into the cognitive-semantic indicators

© Rohovets M., Ievdokymov V., 2026

This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/deed.uk)

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

that parametrize and drive the graph, and that triggers an adaptive response from a detected anomaly rather than from an analyst's manual impulse.

The contribution is threefold: an end-to-end, modular natural-language-processing (NLP) pipeline that maps text to concept-level risk indicators without per-step expert re-estimation; an anomaly-triggered coupling between the detector and the cognitive map, quantified by early-warning lead time; and a reproducible computational experiment reporting pipeline-level metrics.

LITERATURE REVIEW

Research on adaptive and resilience-oriented governance establishes that hybrid threats are nonlinear and cross-domain, and therefore incompatible with rigid hierarchical decision systems [1], [8], [9]. To operationalize adaptability, decision-support systems increasingly incorporate machine learning and predictive analytics, while cautioning that automated tools must complement rather than replace human judgement [10], [11], [12].

From the technology side, cognitive maps provide a mathematically robust mechanism for modelling causal interactions and simulating alternative scenarios [13], [14], and have been applied to cybersecurity and information-risk management [15], [16]. In parallel, text mining and NLP supply the extraction layer: tokenization, TF-IDF weighting, semantic-similarity scoring and contextual clustering convert unstructured text into features [17], [18]. Unsupervised anomaly detection – in particular Isolation Forest – is well suited to streaming, high-dimensional, weakly-labelled signals [19], [20]. Soft-computing methods such as neuro-fuzzy expert systems have likewise been applied to evaluating complex organizational and technical systems [21].

Problem statement. Despite mature components on both sides, the coupling between an NLP extraction layer and a cognitive-semantic decision model is rarely made explicit or reproducible for the security domain, and even less so for Ukrainian-language sources. There is a need for an integrated, portable pipeline that closes this loop and reports its behaviour quantitatively.

METHODS

Pipeline architecture. The system is organized as five sequential stages (Fig. 1). The input is a stream of short Ukrainian-language documents; the output is the integral state-security indicator $C_7(t)$ produced by a downstream cognitive map. Each stage is independently replaceable, allowing the lexical components used here to be swapped for transformer embeddings without changing the interfaces.

S1 – Ingestion and preprocessing. Documents are tokenized with a Cyrillic/Latin token pattern (tokens of length ≥ 3) and normalized. Each record carries a time bucket t (one calendar day in this study) and its text.

S2 – Concept mapping. Seven security concepts are used, adopted from the prior model [7]: C_1 Russian aggression (military/cyber); C_2 critical infrastructure; C_3 institutional capacity; C_4 social resilience; C_5 information space; C_6 international support; C_7 state security (goal). For each concept a TF-IDF anchor vector is built as the mean of a small set of labelled seed texts; an incoming document is assigned to the concept of maximal cosine similarity. On the demonstration corpus this reproduced the reference labelling with 89.1% agreement (Fig. 5).

S3 – Indicator scoring. For each concept c and day t a semantic risk indicator is computed as the mean destructive polarity of that concept's documents, scaled by their volume:

$$R_c(t) = \bar{p}_c(t) \cdot \ln(1 + v_c(t)), \quad (1)$$

where $\bar{p}_c(t) \in [-1, 1]$ is the mean polarity from a destructive/stabilising lexicon (positive = escalatory, negative = stabilising) and $v_c(t)$ is the document count. Threat concepts (C_1, C_2, C_5) contribute positively to aggregate risk; support concepts (C_3, C_4, C_6) contribute negatively (Fig. 2).



Fig. 1. **Five-stage architecture of the cognitive-semantic indicator pipeline**
***Source*: compiled by the authors**

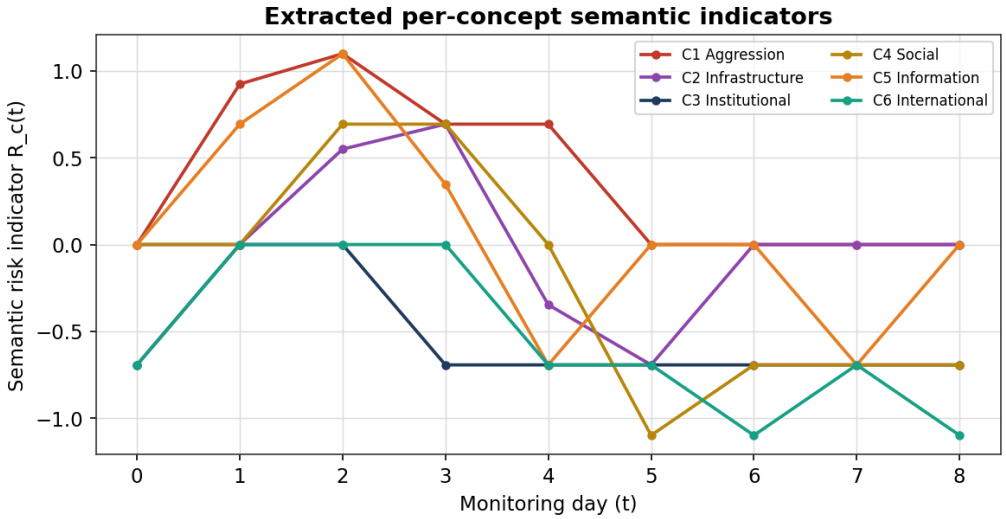


Fig. 2. **Extracted per-concept semantic risk indicators over the monitoring window**
***Source*: compiled by the authors**

S4 – Anomaly detection. The aggregate net risk R(t) is monitored by a one-sided sliding-window z-score with a two-day warmup, which serves as the operational threat trigger:

$$z_t(t) = (R_t - \mu_t) / \sigma_t, \quad flag\ if\ z_t > k, \qquad (2)$$

with μ_t and σ_t estimated from the preceding window and threshold k = 1.5. An Isolation Forest (200 estimators) over the feature vector [R(t), threat, support] provides two-sided corroboration. Because Isolation Forest also flags the recovery trough (an excess of stabilising signal), the one-sided z-score is used as the threat trigger, while the forest supplies structural confirmation (Fig. 3).

S5 – Driving the cognitive map. The extracted indicators drive a weighted cognitive map over the seven concepts, whose adjacency matrix W is adopted from [7]. Each concept holds a health state $S_i(t) \in [0, 1]$ (1 = healthy). C_1 (aggression) is set exogenously from its extracted indicator; the remaining concepts evolve as a deviation-driven update with natural relaxation toward a healthy baseline b_i:

$$S_i(t+1) = \text{clip}[\, S_i(t) + \eta \Sigma_j w_{ji} (S_j(t) - b_j) + \\ +\rho(b_i - S_i(t)) + \sigma_i(t) + a_i(t) \,], \qquad (3)$$

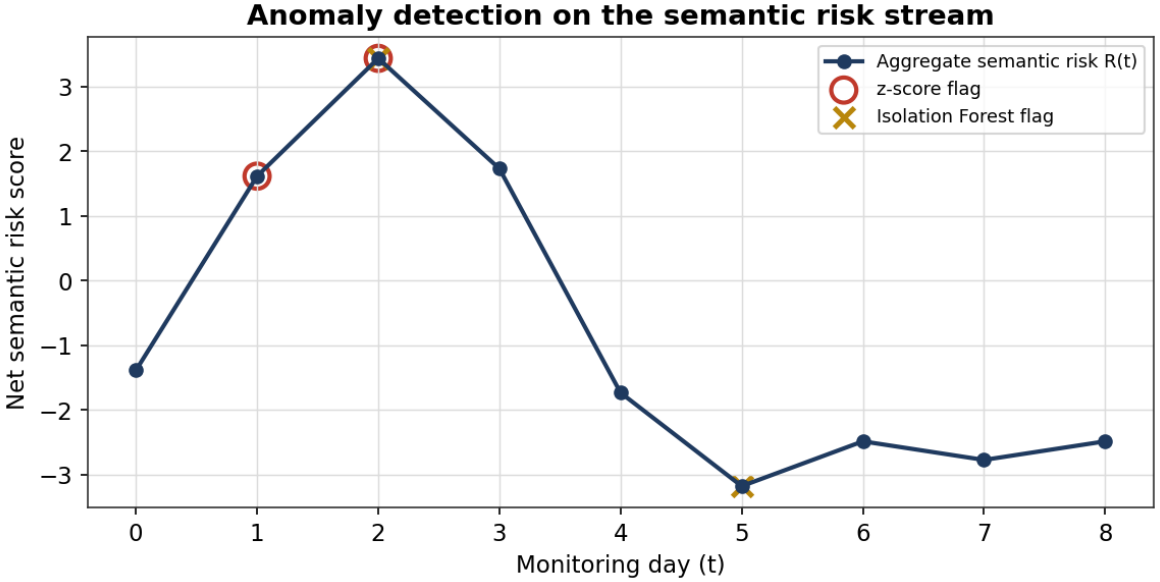


Fig. 3. **Anomaly detection on the aggregate semantic risk stream**
***Source*: compiled by the authors**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

where η is the propagation gain, ρ the relaxation rate, $\sigma_i(t)$ exogenous shocks from S3, and $a_i(t)$ the adaptive-governance response. In the adaptive branch, once S4 raises an anomaly a counter-impulse is engaged on the institutional (C_3), international (C_6) and strategic-communication (C_5) nodes and sustained with decay; in the baseline branch $a_i(t) = 0$.

RESEARCH RESULTS

The pipeline was exercised end-to-end on a structured demonstration corpus (46 records, nine daily buckets, with a combined-escalation event injected at t_2–t_3). Table 1 summarizes the stages; Table 2 reports the run.

Early warning. The detector raised its first flag at t_1, two days before the modelled escalation peak at t_3, and coincided with the known onset at t_2. In an operational setting this lead time is the window available for pre-emptive governance action.

State-security dynamics. Driven by the extracted indicators, the cognitive map reproduced a sharp divergence between the two governance regimes (Fig. 4). Without adaptive response the integral indicator C_7 declined monotonically to ≈0.05 of the normative level – an effectively irreversible collapse. With the anomaly-triggered adaptive response, the decline was arrested at ≈0.56 and the system recovered to a stable plateau of ≈0.61. Crucially, unlike the prior model, neither the driving impulse nor the moment of intervention was set by hand: both were produced by the pipeline from the text stream.

DISCUSSION OF RESULTS

What the automation adds. The prior model demonstrated that adaptive interventions on institutional capacity and international support stabilize state security [7], building on the authors' established link between state security and national resilience under hybrid threats [22]. The present result reproduces that qualitative behaviour but removes the manual step: the impulse magnitude and timing are derived from measured semantic dynamics. This converts a static, expert-parametrized construct into a continuously updatable monitoring instrument – the practically decisive difference for real-time use.

Detector directionality. The two-sided Isolation Forest flags both the escalation and the recovery trough; only the former is a threat. Using a one-sided z-score as the operational trigger and the forest as structural corroboration resolves this and avoids acting on benign over-stabilisation – a design point that generalizes to any bidirectional anomaly signal feeding a decision loop.

Table 1. **Stages of the cognitive-semantic indicator pipeline**

Stage	Function	Method	Output
S1	Ingestion & preprocessing	Ukrainian tokenization, normalization	Clean token stream
S2	Concept mapping	TF-IDF anchors + cosine similarity	Document → $C_1 \ldots C_7$
S3	Indicator scoring	Polarity lexicon × volume (Eq. 1)	$R^c(t)$ per concept/day
S4	Anomaly detection	Sliding z-score (Eq. 2) + Isolation Forest	Threat flags + lead time
S5	FCM driving	Deviation-FCM update (Eq. 3)	$C_7(t)$ trajectory

***Source*: compiled by the authors**

Table 2. **Pipeline run on the demonstration corpus**

Metric	Value
Records processed	46
Daily buckets	9
Concept-mapping agreement (vs seed labels)	89.1 %
Known event onset	t_2
First anomaly flag	t_1 (2 days before peak t_3)
Sliding z-score flags	t_1, t_2
Isolation Forest flags	t_2, t_5
C_7 baseline minimum	0.05 (≈95% deviation)
C_7 adaptive minimum	0.56 (≈44% deviation)
C_7 adaptive plateau (end)	0.61

***Source*: compiled by the authors**

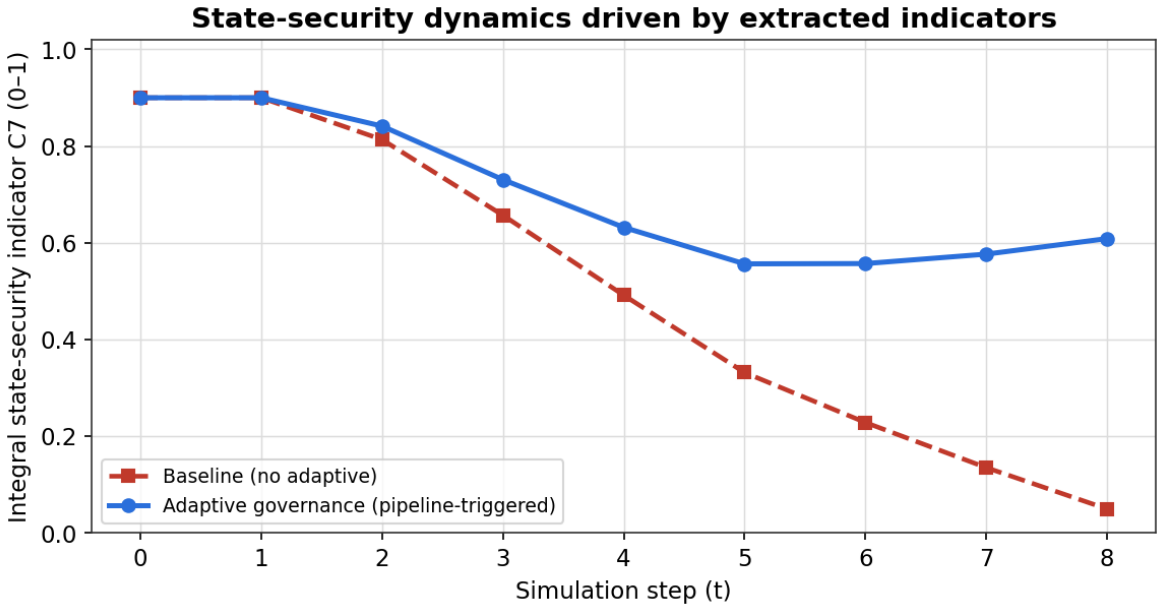


Fig. 4. **State-security dynamics $C_7(t)$: baseline vs. anomaly-triggered adaptive governance**
***Source*: compiled by the authors**

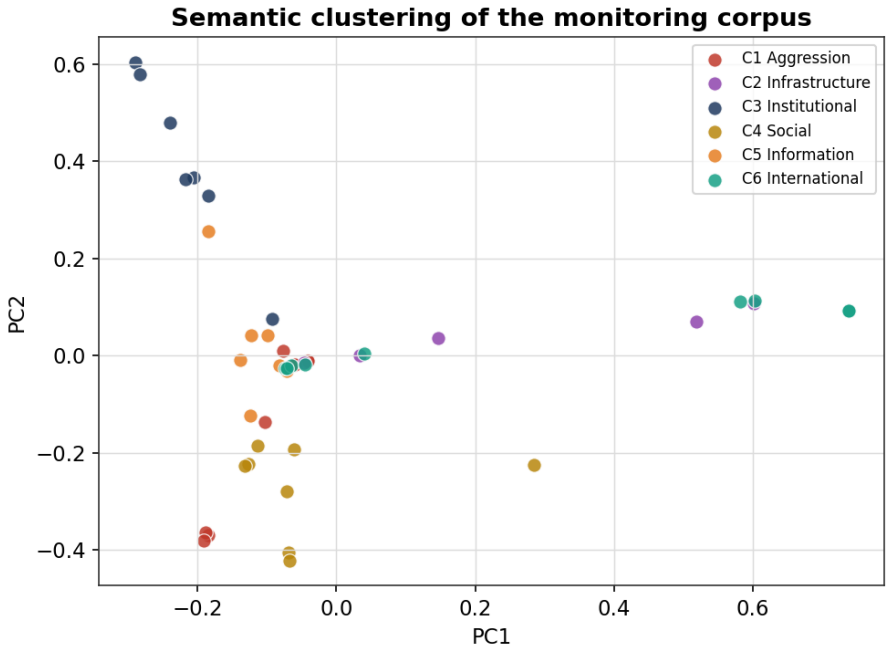


Fig. 5. **Semantic clustering of the monitoring corpus**
(PCA of TF-IDF features, coloured by predicted concept)
***Source*: compiled by the authors**

Limitations. The reported figures come from an illustrative demonstration corpus authored to exercise the pipeline end-to-end; they are proof-of-concept, not operational measurements. The corpus is small and its escalation event is injected by design. Polarity is lexicon-based and therefore sensitive to vocabulary coverage; the FCM weights are inherited from expert elicitation. These do not affect the architecture or its reproducibility, but they bound the numeric claims until the pipeline is run on an operational stream.

CONCLUSIONS

This study contributed an end-to-end NLP pipeline that extracts cognitive-semantic risk indicators from a Ukrainian-language text stream, detects semantic anomalies, and drives a downstream cognitive map to produce an integral

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

state-security indicator. On a demonstration corpus the pipeline mapped documents to concepts with 89.1% agreement, raised an early-warning flag two days before the escalation peak, and – driven entirely by the extracted signals – reproduced the divergence between an uncontrolled collapse ($C_7 \approx 0.05$) and an anomaly-triggered adaptive recovery to a stable plateau ($C_7 \approx 0.61$).

The scientific contribution is the explicit, reproducible coupling of an NLP extraction layer to a cognitive-semantic decision model, with quantified early-warning behaviour. Practically, the pipeline is modular and portable to an operational monitoring stream. Future work concerns replacing lexical polarity with transformer embeddings, validating on an operational corpus, and learning the adjacency weights from data rather than expert elicitation.

Conflict of interests: Authors declare that they do not have conflict of interests, in particular financial, personal, copyright or any of a different nature, which could have influenced the research, as well as the results published in this articles

Financing: Research was carried out without financial support

Accessibility data: All data available in digital or graphic form in the main text manuscript

Using artificial intelligence tools: During the preparation of this manuscript the authors used a large language model (AI assistant) to support language editing and drafting of the text, the implementation of the demonstration pipeline code, and the generation of the figures from the computed results. All AI-assisted outputs were reviewed, verified and edited by the authors. The research design, scientific provisions, interpretation of the results and conclusions are the authors' own intellectual contribution, and the authors take full responsibility for the content of the article.

REFERENCES

1. Žilinskas, R. "National resilience against external hybrid threats: A hypothetical model". *Politologija*. 2017; 87 (3): 45–87. DOI: https://doi.org/10.15388/Polit.2017.3.10856.

2. Hartmann, S. "The role of science in resilience planning for military-civilian domains in the U.S. and NATO". *Defense & Security Analysis*. 2024; 40 (3): 341–360. DOI: https://doi.org/10.1080/14702436.2024.2365218.

3. Mykhailova, Y. V. "European integration as a mechanism for strengthening economic resilience". *Scientific Innovations and Advanced Technologies.* 2025; 7 (47): 538–547. DOI: https://doi.org/10.52058/2786-5274-2025-7(47)-538-547.

4. World Economic Forum. "Emerging paradigms in security crisis management: Technology, collaboration, and adaptive governance". *WEF Reports*. 2025. – Available from: https://www.weforum.org. – [Accessed: March 2026].

5. Gunter, C. & Bourgon, J. "Resilience and the new synthesis of public administration". *Public Governance International Working Paper*. 2019. – Available at: https://www.pgionline.com. – [Accessed: March 2026].

6. Abdillah, A., Widianingsih, I., Buchari, R. A. & Nurasa, H. "Resilience-based governance: A public administration perspective and resilience agenda". *Public Organization Review.* 2025; 25: 695–715. DOI: https://doi.org/10.1007/s11115-025-00834-z.

7. Rohovets, M. & Ievdokymov, V. "Methodology for assessing the state of the security environment and the level of national security of the state based on a cognitive-semantic model" [In Ukrainian]. *Society and Security.* 2025; 5 (11): 73–81. DOI: https://doi.org/10.26642/sas-2025-5(11)-73-81.

8. He, J., He, Y., Hu, J. & Guo, Y. "Intelligent governance: The AI-driven new paradigm of governmental adaptive governance". *Journal of US-China Public Administration.* 2025; 22 (1): 1–27. DOI: https://doi.org/10.17265/1548-6591/2025.01.001.

9. Žilinskas, R. & Sakalauskas, L. "Exploring crisis governance: Quest for functional resilience during COVID-19 in Lithuania". *Scientific Journal of the Military University of Land Forces*. 2021; 53 (4 (202)): 763–798. DOI: https://doi.org/10.5604/01.3001.0015.6180.

10. Dunn Cavelty, M. & Wiedemar, S. "AI decision support systems on the battlefield". *Center for Security Studies (CSS).* ETH Zürich. 2025. – Available from: https://css.ethz.ch. – [Accessed: March 2026].

11. Bode, I. & Huelss, H. "Computer says 'war': AI and resort-to-force decision making". *Cambridge Forum on AI: Law and Governance.* 2026. DOI: https://doi.org/10.1017/cfa.2025.1.

12. “Artificial intelligence and machine learning in armed conflict: A human-centred approach”. *International Committee of the Red Cross*. 2022.– Available from: https://www.icrc.org. – [Accessed: March 2026].

13. Kosko, B. “Fuzzy cognitive maps”. *International Journal of Man-Machine Studies.* 1986; 24 (1): 65–75. DOI: https://doi.org/10.1016/S0020-7373(86)80040-2.

14. Pires, J. & Marreiros, G. “Using neutrosophic cognitive maps to support group decisions about modeling and analyzing smart port performance”. *Applied Sciences.* 2025; 15 (4): 1981. DOI: https://doi.org/10.3390/app15041981.

15. Baranov, V. V. & Shelupanov, A. A. “Cognitive model for assessing the security of information systems for various purposes”. *Symmetry*. 2022; 14 (12): 2631. DOI: https://doi.org/10.3390/sym14122631.

16. Veksler, V. D. et al. “Simulations in cyber-security: A review of cognitive modeling of network attackers, defenders, and users”. *Frontiers in Psychology.* 2018; 9: 691. DOI: https://doi.org/10.3389/fpsyg.2018.00691.

17. Salton, G. & Buckley, C. “Term-weighting approaches in automatic text retrieval”. *Information Processing & Management.* 1988; 24 (5): 513–523. DOI: https://doi.org/10.1016/0306-4573(88)90021-0.

18. Aggarwal, C. C. & Zhai, C. “A survey of text classification algorithms”. In *Mining Text Data, Springer*. 2012. p. 163–222. DOI: https://doi.org/10.1007/978-1-4614-3223-4_6.

19. Liu, F. T., Ting, K. M. & Zhou, Z.-H. “Isolation Forest”. In *Proc. 8th IEEE Int. Conf. on Data Mining (ICDM)*. 2008. p. 413–422, https://www.scopus.com/pages/publications/67049142378. DOI: https://doi.org/10.1109/ICDM.2008.17.

20. Ding, Z. & Fei, M. “An anomaly detection approach based on Isolation Forest algorithm for streaming data using sliding window”. *IFAC Proceedings Volumes.* 2013; 46 (20): 12–17. DOI: https://doi.org/10.3182/20130902-3-CN-3020.00044.

21. Shyshatskyi, A., et al. “Development of a method for evaluating complex organizational and technical systems using neuro-fuzzy expert systems”. *Eastern-European Journal of Enterprise Technologies*. 2025; 6 (4 (138)): 6–14. DOI: https://doi.org/10.15587/1729-4061.2025.344556.

22. Rohovets, M. A. “Correlation between state security and national resilience under hybrid threats” [In Ukrainian]. *Public Administration: Improvement and Development.* 2026; 1: 11. DOI: https://doi.org/10.32702/2307-2156.2026.1.11.

Received 28.08.2026
Received after revision 22.10.2026
Accepted 30.10.2026

DOI: https://doi.org/10.15276/hait.09.2026.46
UDC 004.85:004.62:351.86

Когнітивно-семантичне витягування індикаторів безпеки в реальному часі: конвеєр обробки природної мови для адаптивного врядування на основі виявлення аномалій

Роговець Максим Анатолійович[1)]
ORCID: https://orcid.org/0000-0002-1587-9017; mrohovets@gmail.com
Євдокимов Віктор Валерійович[2)]
ORCID: https://orcid.org/0000-0002-3577-081X; viktorievdokymov@gmail.com
[1)] Науково-дослідний інститут воєнної розвідки, вул. Юрія Іллєнка, 81. Київ, 04050, Україна
[2)] Державний університет “Житомирська політехніка”, вул. Чуднівська, 103. Житомир, 10005, Україна

АНОТАЦІЯ

Актуальність. Адаптивне врядування безпековим середовищем в умовах гібридної агресії потребує перетворення великих зашумлених текстових потоків на машинно-обчислювані індикатори достатньо швидко, щоб діяти на випередження. Попередні моделі параметризують граф вручну, що не масштабується на моніторинг у реальному часі. **Мета.** Розробити та апробувати наскрізний конвеєр обробки природної мови, який автоматично витягує когнітивно-семантичні

індикатори ризику з текстового потоку, виявляє семантичні аномалії та живить когнітивну карту для обчислення інтегрального показника державної безпеки в реальному часі. **Методи.** Багатостадійний конвеєр: препроцесинг україномовного тексту; віднесення до набору безпекових концептів через якорі на основі міри «частота терміна — обернена документна частота» (term frequency – inverse document frequency, TF-IDF) та косинусну близькість; підрахунок індикаторів (лексикон деструктивної полярності × обсяг); виявлення аномалій ковзним z-показником із підтвердженням методом Isolation Forest; живлення зваженої когнітивної карти (структура суміжності — з попередньої моделі авторів) із адаптивною відповіддю, що вмикається за аномалією. **Результати.** Віднесення до концептів відтворило еталонне маркування з високим рівнем узгодженості. Виявлення аномалій підняло перший сигнал завчасно до піку ескалації, даючи дієве вікно раннього попередження. Під впливом витягнутих індикаторів без адаптивної відповіді інтегральний показник обвалився до критично низької частки нормативного рівня, тоді як адаптивна гілка зупинила падіння й вивела систему на стійке плато. **Висновки.** Автоматизація витягування індикаторів закриває розрив між сирими інформаційними потоками та когнітивною моделлю, перетворюючи статичну конструкцію на інструмент безперервного моніторингу.

Ключові слова: обробка природної мови; текст-майнінг; виявлення аномалій; Isolation Forest; когнітивна карта; безпекове середовище; адаптивне врядування

ABOUT THE AUTHORS



Maksym A. Rohovets - Candidate of Technical Sciences, Associate Professor, Head of Department. Defense Intelligence Research Institute. 81, Yurii Illienko Str. Kyiv, 04050, Ukraine
ORCID: https://orcid.org/0000-0002-1587-9017; mrohovets@gmail.com
***Research field*:** National Security, State Security, Military Security, Artificial Intelligence, Information Technologies

Роговець Максим Анатолійович – кандидат технічних наук, доцент, начальник Управління. Науково-дослідний інститут воєнної розвідки, вул. Юрія Іллєнка, 81. Київ, 04050, Україна
Галузь досліджень: національна безпека, державна безпека, воєнна безпека, штучний інтелект, інформаційні технології



Victor V. Ievdokymov - Doctor of Economic Sciences, Professor, Professor of the Department of National Security, Public Administration and Governance. Zhytomyr Polytechnic State University, 103, Chudnivska Str. Zhytomyr, 10005, Ukraine
ORCID: https://orcid.org/0000-0002-3577-081X; viktorievdokymov@gmail.com
***Research field*:** Accounting, Financial Analysis and Auditing, Public Administration, State Security, Public Order Protection

Євдокимов Віктор Валерійович - доктор економічних наук, професор, професор кафедри Національної безпеки, публічного управління та адміністрування. Державний університет “Житомирська політехніка”, вул. Чуднівська, 103. Житомир, 10005, Україна
Галузь досліджень: облік, фінансовий аналіз та аудит, публічне управління та адміністрування, державна безпека, охорона громадського порядку