BIMA
BULLETIN OF INTELLIGENT MACHINES AND ALGORITHMS
Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.31

A Trigger Aware, Event Centric, and Uncertainty Calibrated Neuro-Symbolic Framework for Actionable Cyber Threat Intelligence from Indonesian Online News

Elia Setiana[1], Muhamad Achya Arifudin[2], Nur Alamsyah[3]

Informatics[1,2], Information System[3]
Universitas Informatika dan Bisnis Indonesia, Bandung, Indonesia[1,2,3]
https://unibi.ac.id[1,2,3]
elia.setiana@unibi.ac.id[1], achyarifudin@unibi.ac.id[2], nuralamsyah@unibi.ac.id[3]

Abstract. Online news can provide timely cyberthreat signals, but duplicative reporting, fragmented event descriptions, resource-constrained Indonesian language text, and uncalibrated model confidence limit its operational use. This study presents A Trigger-Aware, Event-Centric, and Uncertainty-Calibrated Neuro-Symbolic Framework for Actionable Cyber Threat Intelligence from Indonesian Online News (TRACE-CTI-ID), a proof-of-concept framework that integrates exact deduplication, event-centric clustering, trigger-aware semantic representation, neuro-symbolic fusion, ordinal risk estimation, conformal uncertainty, mitigation mapping, and an event-centric knowledge graph. The experiment used 711 Liputan6 records collected on March 14, 2025. Exact deduplication reduced the corpus to 79 unique headlines, which were automatically consolidated into 26 events. Splitting the separate events resulted in 30 training articles, 5 calibration articles, and 44 test articles with zero event leakage. The calibrated neuro-symbolic model achieved a micro-F1 of 0.283 and a macro-F1 of 0.441, outperforming the baseline TF-IDF of 0.074 and 0.013, respectively. However, ordinal severity prediction remained weak with an accuracy of 0.136, a macro-F1 of 0.138, and a mean absolute error of 2.023. Conformal coverage was also unstable, and the abstention mechanism did not direct uncertain articles to human review. These findings demonstrate the technical feasibility of the integrated pipeline while also demonstrating that silver labeled, title only, and single source data are insufficient for final operational validation. Therefore, the key contribution is a transparent, leak aware evaluation architecture and protocol that can be strengthened through full text collection from multiple sources and independent expert annotation.

Key words: cyber threat intelligence; Indonesian news; event clustering; trigger-aware learning; conformal prediction

1. INTRODUCTION

Cyber threat intelligence (CTI) transforms observations about adversaries, infrastructure, behavior, and impact into knowledge that supports detection, prioritization, and response. Structured standards such as STIX 2.1 provide a machine-readable representation for sharing CTI [1], while MITRE ATT&CK organizes adversary behavior into tactics, techniques, and sub-techniques [2]. NIST Cybersecurity Framework 2.0 complements these resources by framing cybersecurity outcomes around governance, identification, protection, detection, response, and recovery [3]. In practice, however, relevant evidence is often first disclosed through unstructured reports, vendor blogs, official announcements, and online news. Security analysts must therefore identify useful signals from a rapidly growing and repetitive text stream before the information can be operationalized.

Existing NLP-based CTI research has progressed from indicator-of-compromise extraction toward richer contextual intelligence. TriCTI introduced campaign triggers to associate indicators with attack stages [4]. TIM combined contextual descriptions and security elements to classify ATT&CK-oriented tactics, techniques, and procedures [5]. Vulcan extracted threat entities and semantic relations and stored them in a graph database for profiling and evolution analysis [6]. Silvestri et al. linked NLP-based threat and asset extraction to severity assessment and mitigation in healthcare ecosystems [7], while KnowCTI incorporated cybersecurity knowledge into entity and relation extraction to generate a threat-intelligence graph [8]. These studies establish that domain

Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.31

knowledge and contextual features can improve the transformation of unstructured text into structured intelligence.

Three gaps remain important for Indonesian online news. First, document-centric processing can over-count a single incident when many articles report the same event. This problem is visible in scraped news datasets where duplicate records and semantically similar headlines are common. Second, most CTI extraction systems issue deterministic predictions without explicitly indicating when a result is unreliable. Temperature scaling can reduce overconfidence [9], and conformal prediction offers model-agnostic prediction sets with statistical coverage objectives [10], but uncertainty-aware CTI from Indonesian news remains underexplored. Third, Indonesian is a comparatively low-resource NLP setting. IndoBERT and the IndoLEM and IndoNLU benchmarks have improved Indonesian language processing [11], [12], while multilingual models such as XLM-R [13] and Sentence-BERT [14] offer practical representations for cross-lingual and semantic similarity tasks. Nevertheless, a complete event-aware, knowledge-guided, and uncertainty-calibrated CTI pipeline has not been established for Indonesian cyber-news.

This paper proposes TRACE-CTI-ID, a trigger-aware, event-centric, and uncertainty-calibrated neuro-symbolic framework for actionable CTI from Indonesian online news. The framework integrates data-quality control, event consolidation, semantic and trigger features, symbolic cyber rules, ordinal risk classification, confidence calibration, conformal prediction, mitigation mapping, and knowledge-graph construction. The present experiment is deliberately positioned as a proof of concept because it uses a small, title-only, single-source corpus and silver labels produced through weak supervision. The paper contributes: (1) an end-to-end architecture that connects event deduplication to actionable CTI; (2) an event-disjoint evaluation protocol designed to prevent event leakage; (3) a transparent comparison among TF-IDF, trigger-aware, and neuro-symbolic variants; and (4) an explicit analysis of failure modes in clustering, severity estimation, calibration, and abstention.

2. RELATED WORK

2.1. NLP-based cyber threat intelligence

Automated CTI initially emphasized indicators such as IP addresses, domains, URLs, hashes, and vulnerability identifiers. Although these indicators are useful for blocking known infrastructure, they are short-lived and often insufficient to describe adversary behavior. TriCTI addressed this limitation by introducing campaign-trigger phrases that emphasize the words most indicative of an attack stage [4]. TIM expanded the context by combining textual descriptions with twelve categories of TTP elements and produced STIX-formatted intelligence [5]. Both approaches demonstrate that explicit trigger and context features can improve task-specific interpretation beyond generic text classification.

Entity and relation extraction further enrich CTI. Vulcan used language-model-based named entity recognition and relation extraction to identify descriptive threat information and support graph-based analysis [6]. KnowCTI injected external cybersecurity triples through attention and graph-based knowledge fusion, reporting F1 scores above 90 for entity extraction and above 81 for relation extraction [8]. Silvestri et al. connected extracted threats and assets to threat prioritization and mitigation selection in healthcare [7]. These approaches are powerful, but they generally assume that each document is processed as an independent report and do not explicitly address duplicated or cross-document news coverage.

2.2. Indonesian representation and event consolidation

IndoLEM and IndoBERT established reusable Indonesian resources for morpho-syntactic, semantic, and discourse tasks [11], while IndoNLU provided twelve Indonesian natural-language-understanding tasks and IndoBERT models trained on the Indo4B corpus [12]. XLM-R supports multilingual transfer at scale [13]. Sentence-BERT produces semantically meaningful sentence vectors suitable for similarity search and clustering [14]. These resources motivate semantic clustering for Indonesian news, but event consolidation requires more than sentence similarity: articles should also agree on time, actor, target, and threat type.

Uncertainty is equally important because CTI decisions may trigger incident-response actions. Guo et al. showed that modern neural models are frequently miscalibrated and that temperature scaling is an effective post-hoc method [9]. Campos et al. reviewed conformal prediction for NLP and highlighted its distribution-free, model-agnostic value for uncertainty quantification [10]. TRACE-CTI-ID combines these ideas with human review, but the present study also demonstrates that conformal guarantees are not meaningful when calibration samples are too few or nonconformity logic is incorrectly designed, as shown in Table 1.

TABLE 1. Positioning Of Trace CTI ID Against Representative CTI Studies

Study	Main contribution	Gap addressed by TRACE-CTI-ID
TriCTI [4]	Trigger-enhanced campaign-stage classification	Adds event consolidation, severity, uncertainty, and Indonesian news.
TIM [5]	Context-enhanced TTP mining and STIX output	Adds duplicate-event control and explicit abstention.
Vulcan [6]	Entity/relation extraction and graph applications	Adds event-centric cross-document processing.

Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.31

Study	Main contribution	Gap addressed by TRACE-CTI-ID
Silvestri et al. [7]	Threat assessment and mitigation for healthcare	Generalizes toward Indonesian news and uncertainty-aware decisions.
KnowCTI [8]	Knowledge-enhanced entity/relation extraction	Adds event/time/source-disjoint validation and ordinal severity.

3. METHODS

This study adopts a proof-of-concept experimental design to evaluate TRACE-CTI-ID as an integrated pipeline for transforming Indonesian cybersecurity news into structured and uncertainty-aware cyber threat intelligence. The methodological workflow covers five stages: data preparation and exact deduplication; event-centric consolidation and event-disjoint grouping; trigger-aware semantic representation with neuro-symbolic threat classification; ordinal severity estimation with calibration, conformal uncertainty, mitigation mapping, and knowledge-graph construction; and leakage-safe evaluation. Because the corpus is limited to headline-only, single-source records with silver labels, the experiment is intended to assess technical feasibility and failure modes rather than final operational accuracy. Figure 1 summarizes the overall workflow, while Sections 3.1-3.5 detail each stage.

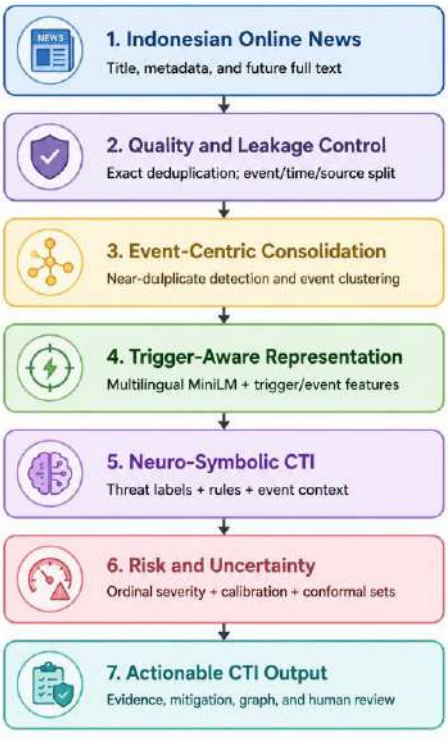

Fig. 1. Proposed TRACE-CTI-ID workflow.

3.1. Research design and dataset

The experiment used the spreadsheet Liputan6_Scraping_keamanan siber.xlsx. It contained four attributes: news title, relative publication time, source, and scraping date. All records came from Liputan6 and were scraped on 14 March 2025. Because full article text and URLs were unavailable, the title was used as the textual input. Relative dates such as "six months ago" were converted into approximate dates anchored to the scraping date. The experiment therefore evaluates technical pipeline feasibility rather than complete CTI extraction from full reports.

Figure 2 shows that quality control removed exact duplicate titles and generated article identifiers. The raw dataset contained 711 rows; exact deduplication retained 79 unique articles and removed 632 duplicated rows. Silver multi-label categories were then assigned through weak supervision using cyber-threat dictionaries, trigger rules, and title patterns. The nine labels were general cybersecurity, AI/deepfake, awareness and capacity, cyber-business solution, policy resilience, ransomware, intrusion/hacking, phishing/fraud, and data breach/privacy. Severity labels were similarly generated on an ordinal scale: Informational, Low, Medium, High, and Critical, as shown in Table 2.

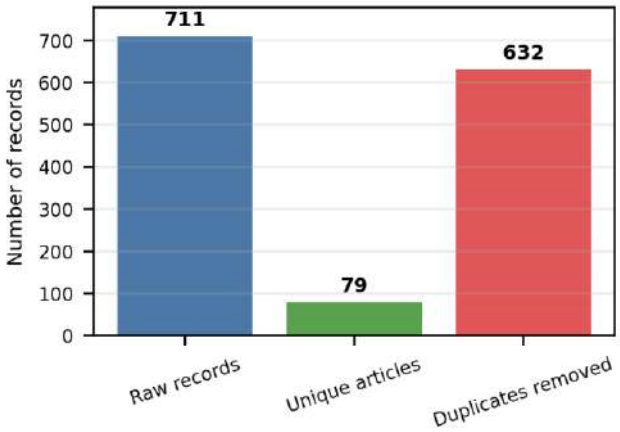

Fig. 2. Raw records, unique articles, and exact duplicates removed.

TABLE 2. Experimental Dataset Profile

Item	Value
Raw / unique / removed	711 / 79 / 632 (88.89%)
Source and text unit	Liputan6; headline only
Automatic events	26
Train / calibration / test	30 / 5 / 44
Label source	Silver labels from weak supervision

3.2. Event-centric consolidation

Each headline was encoded using paraphrase-multilingual-MiniLM-L12-v2, a multilingual sentence-transformer derived from the sentence-embedding paradigm [14]. Pairwise semantic similarity supported agglomerative clustering. The intended event score combines semantic similarity, entity overlap, temporal proximity, target agreement, and threat agreement:

$$S_{event} = w1\ S_{text} + w2\ S_{entity} + w3\ S_{time} + w4\ S_{target} + w5\ S_{threat} \quad (1)$$

Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.31

In the current prototype, semantic similarity dominated because reliable entities, full text, and exact publication dates were unavailable. The procedure formed 26 automatic events, 12 of which contained more than one article. The largest cluster contained 26 headlines, revealing a risk of over-merging generic cybersecurity stories. Event IDs were used as group identifiers for leakage-safe partitioning.

3.3. Trigger-aware and neuro-symbolic threat classification

The baseline used TF-IDF features and one-vs-rest logistic regression. The trigger-aware representation concatenated multilingual sentence embeddings with binary trigger features, event-size information, and weakly extracted cyber concepts. Trigger lexicons represented attack, compromise, response, and protection expressions such as serangan, peretasan, ransomware, pencurian data, melindungi, and memperkuat. The conceptual multi-task objective is:

$$\mathcal{L}(\text{MTL}) = \lambda_1\mathcal{L}(\text{threat}) + \lambda_2\mathcal{L}(\text{trigger}) + \lambda_3\mathcal{L}(\text{entity}) + \lambda_4\mathcal{L}(\text{relation} \quad (2)$$

The prototype did not fine-tune a large Transformer because only 79 unique headlines were available. Instead, it trained light-weight classifiers on pretrained semantic representations. Symbolic rules mapped recognized patterns to threat categories, impacts, and mitigation actions. Neural probabilities were fused with symbolic consistency scores:

$$P_{ns}(y|x,G) = \text{Normalize}[P_n(y|x)\exp(\beta R(y,G))] \quad (3)$$

Here, P_N denotes the learned probability, R(y,G) is a rule and graph consistency score, and beta controls symbolic influence. The design reflects ATT&CK concepts, NIST CSF outcomes, and STIX-style entities and relations [1]-[3].

3.4. Severity, calibration, and conformal prediction

Severity was modeled as an ordered target rather than an unordered class. The prototype learned severity from semantic, trigger, threat, and symbolic features. Temperature scaling optimized a scalar T on the calibration set to transform logits z into calibrated probabilities softmax(z/T) [9]. Conformal prediction then generated a set of plausible labels:

$$\Gamma_a(x) = \{y : s(x,y) \le q_{(1-a)}\} \quad (4)$$

The target coverage was 90%. Cases with multiple labels, empty sets, conflicting evidence, or low confidence should be routed to a human analyst. Evidence-grounded mitigation was produced by mapping predicted threats to response actions, including incident-response activation, credential rotation, IOC validation, phishing education, backup verification, and security monitoring. Extracted event, threat, actor, target, impact, and mitigation nodes were stored in an event-centric graph.

3.5. Evaluation protocol

GroupShuffleSplit partitioned the data by event, producing 30 training, 5 calibration, and 44 test headlines. No event appeared in more than one split. Multi-label threat classification was evaluated using micro-F1, macro-F1, micro-precision, micro-recall, and Hamming loss. Ordinal severity was evaluated using accuracy, macro-F1, mean absolute error (MAE), and quadratic weighted kappa (QWK). Calibration was inspected through empirical coverage and average prediction-set size. Because no gold event IDs or expert threat annotations were available, the reported metrics validate internal pipeline behavior against silver labels and must not be interpreted as final real-world accuracy.

4. RESULTS AND DISCUSSIONS

4.1. Data quality and event leakage

Exact deduplication removed 632 of 711 rows, equivalent to an 88.89% duplication rate. This result confirms that data-quality auditing is not a peripheral preprocessing step: random splitting before deduplication would place identical headlines in both training and testing sets and inflate reported performance. The event-disjoint split achieved zero shared event IDs across training, calibration, and testing. This is a positive engineering result and directly supports the event-centric novelty of TRACE-CTI-ID.

The clustering result, however, was not yet reliable. The method produced 26 events from 79 headlines, with the largest event containing 26 articles. Many of those titles shared generic expressions such as “keamanan siber” but did not necessarily describe one real-world incident. This over-merging shows that semantic headline similarity alone is insufficient. Future clustering must require temporal compatibility and shared actors, targets, or threat types, followed by expert event annotation and B-cubed F1, adjusted Rand index, and normalized mutual information evaluation.

4.2. Threat classification

TABLE 3. Multi-label threat classification results

Model	Micro-F1	Macro-F1	Precision	Recall	Hamming
TF-IDF logistic	0.074	0.013	0.333	0.042	0.126
Trigger-aware neural	0.241	0.273	0.188	0.333	0.255
Neuro-symbolic calibrated	0.283	0.441	0.228	0.375	0.230

Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.31

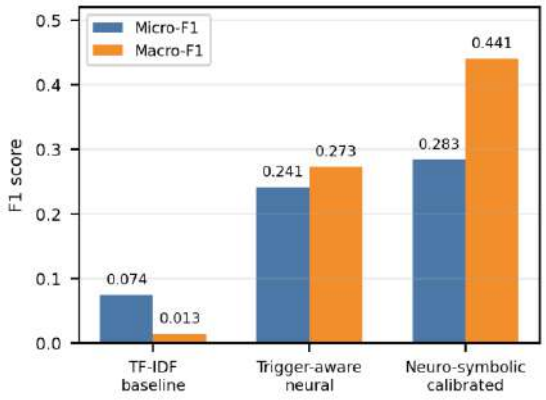

Fig. 3. Micro-F1 and macro-F1 comparison among experimental variants.

Figure 3 shows that the trigger-aware model improved micro-F1 from 0.074 to 0.241 and macro-F1 from 0.013 to 0.273. The calibrated neuro-symbolic variant obtained the best micro-F1 of 0.283 and macro-F1 of 0.441. Its macro-F1 gain is especially relevant because the label distribution was highly imbalanced: general cybersecurity had 38 positive examples, whereas data breach/privacy had only one. Symbolic rules and trigger cues therefore helped recover some minority-label behavior that TF-IDF did not capture.

The absolute performance remained low. The neuro-symbolic model reached only 0.228 precision and 0.375 recall, indicating both false alarms and missed labels. The baseline recorded a lower Hamming loss of 0.126, but its recall was only 0.042; this indicates a conservative classifier that predicted few positive labels rather than a useful detector. The proposed model made more positive decisions, increasing recall and macro-F1 at the cost of additional label errors, as shown in Table 3.

The apparent improvement must also be interpreted under the circular-labeling threat. Silver labels were created using keyword rules, and related keyword features were subsequently supplied to the classifier and symbolic fusion layer. Consequently, the evaluation is not independent: the model partly learns the same rules that created the test labels. A publishable confirmatory experiment must replace test labels with independent expert annotations while retaining symbolic knowledge only as a model input.

4.3. Severity and uncertainty calibration

Table 4 show that severity estimation was the weakest predictive component. Accuracy was 0.136, macro-F1 was 0.138, MAE was 2.023, and QWK was 0.057. These values show that headlines rarely provide sufficient evidence for reliable risk grading. A title may mention ransomware or data theft without specifying affected assets, operational disruption, record counts, exploitability, or recovery status. Severity should therefore be derived from full article evidence and a documented rubric rather than from headlines and weak rules.

TABLE 4. Severity And Conformal Evaluation

Measure	Result	Interpretation
Severity accuracy	0.136	Only a small fraction of test headlines received the exact ordinal class.
Severity macro-F1	0.138	Performance was weak across severity classes.
MAE	2.023	Predictions differed by approximately two severity levels on average.
QWK	0.057	Agreement was only marginally above chance.
Severity conformal coverage	0.136	Far below the 0.90 target.
Average severity set size	1.0	The system did not express multi-class uncertainty.

Conformal prediction also failed to provide the intended reliability guarantee. Threat-label coverage ranged from 0.295 for AI/deepfake and 0.318 for ransomware to 1.0 for several mostly negative labels, while all average set sizes were 1.0. Severity coverage was only 0.136 against the 0.90 target. The calibration set contained only five articles, which is inadequate for nine multi-label targets and five severity classes. In addition, the prototype forced an empty conformal set into a single class, preventing abstention. This design explains why zero articles were routed to human review even when coverage was poor. Empty or oversized sets should instead trigger analyst review.

These negative results are scientifically useful. They demonstrate that adding “conformal prediction” to a pipeline does not itself make the system reliable; the calibration sample, exchangeability assumptions, nonconformity score, label-conditional strategy, and abstention logic must all be validated. Future experiments should use repeated event-grouped cross-validation, a substantially larger calibration partition, Mondrian or label-conditional conformal prediction, and explicit risk-coverage curves.

4.4. Knowledge graph and actionable output

The pipeline generated an event-centric graph containing 72 nodes and 192 edges. Node types represented events, threats, actors, targets, impacts, and mitigations. This demonstrates that the model output can be converted into a structured CTI representation rather than remaining a flat classification result. The graph can support queries such as which events involve ransomware, which organizations are recurrent targets, and which mitigations are linked to a threat category. However, graph quality inherits extraction and clustering errors. A merged event can incorrectly connect unrelated actors, targets, and mitigations, so node and relation accuracy must be verified before operational use.

Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.31

4.5. Threats to validity and publication positioning

Internal validity is limited by circular silver labeling and the use of the same lexical knowledge for label construction and model features. Construct validity is limited because a headline does not contain the full attack context required for actor, technique, impact, and mitigation assessment. External validity is limited by a single media source and a narrow time span. Statistical conclusion validity is limited by 79 unique headlines, 30 training samples, and only five calibration samples.

Accordingly, this study should be reported as a preliminary or proof-of-concept evaluation. Its defensible contribution is the integrated architecture, transparent failure analysis, and event-leakage control—not a claim of operational superiority. A stronger journal experiment requires at least several hundred expert-annotated full-text articles for a pilot and preferably 1,000-3,000 multi-source articles for robust temporal and cross-source evaluation. Two or more annotators should label event identity, threat, actor, target, technique, impact, mitigation, and severity, with inter-annotator agreement reported.

5. CONCLUSIONS

TRACE-CTI-ID was implemented as an end-to-end proof-of-concept framework for transforming Indonesian cybersecurity news into structured and potentially actionable intelligence. The experiment showed that exact deduplication and event-disjoint splitting can eliminate obvious duplicate and event leakage. Trigger-aware semantic features and neuro-symbolic fusion improved micro-F1 and macro-F1 over a TF-IDF baseline, and the pipeline successfully generated an event-centric knowledge graph. These results support the technical feasibility of integrating trigger awareness, event consolidation, symbolic knowledge, ordinal risk, and uncertainty within one workflow.

The study also identified critical limitations. Semantic clustering over-merged generic headlines, severity prediction was weak, conformal coverage did not approach its target, and the abstention mechanism failed to route uncertain cases to human review. Most importantly, silver labels created from lexical rules do not provide an independent evaluation. Therefore, the current results must not be presented as final model accuracy. The next stage is to collect multi-source full-text Indonesian cyber-news, establish expert gold annotations and event identities, redesign label-conditional conformal prediction, and evaluate using repeated event-, time-, and source-disjoint protocols. Under those conditions, TRACE-CTI-ID can be tested as a trustworthy decision-support framework rather than only a functioning prototype.

REFERENCES

[1] B. Jordan, R. Piazza, and T. Darley, eds., "STIX Version 2.1," OASIS Standard, Jun. 2021.

[2] MITRE, "MITRE ATT&CK: A knowledge base of adversary tactics and techniques," 2026. [Online]. Available: https://attack.mitre.org/. Accessed: Jul. 22, 2026.

[3] C. Pascoe, S. Quinn, and K. Scarfone, "The NIST Cybersecurity Framework (CSF) 2.0," NIST CSWP 29, Feb. 2024, doi: 10.6028/NIST.CSWP.29.

[4] J. Liu, J. Yan, J. Jiang, Y. He, X. Wang, Z. Jiang, P. Yang, and N. Li, "TriCTI: An actionable cyber threat intelligence discovery system via trigger-enhanced neural network," Cybersecurity, vol. 5, no. 1, Art. no. 8, 2022, doi: 10.1186/s42400-022-00110-3.

[5] Y. You, J. Jiang, Z. Jiang, P. Yang, B. Liu, H. Feng, X. Wang, and N. Li, "TIM: Threat context-enhanced TTP intelligence mining on unstructured threat data," Cybersecurity, vol. 5, Art. no. 3, 2022, doi: 10.1186/s42400-021-00106-5.

[6] H. Jo, Y. Lee, and S. Shin, "Vulcan: Automatic extraction and analysis of cyber threat intelligence from unstructured text," Computers & Security, vol. 120, Art. no. 102763, Sep. 2022, doi: 10.1016/j.cose.2022.102763.

[7] S. Silvestri, S. Islam, D. Amelin, G. Weiler, S. Papastergiou, et al., "Cyber threat assessment and management for securing healthcare ecosystems using natural language processing," International Journal of Information Security, vol. 23, pp. 31-50, 2024, doi: 10.1007/s10207-023-00769-w.

[8] G. Wang, P. Liu, J. Huang, H. Bin, X. Wang, and H. Zhu, "KnowCTI: Knowledge-based cyber threat intelligence entity and relation extraction," Computers & Security, vol. 141, Art. no. 103824, Jun. 2024, doi: 10.1016/j.cose.2024.103824.

[9] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in Proc. 34th Int. Conf. Machine Learning, PMLR, vol. 70, 2017, pp. 1321-1330.

[10] M. Campos, A. Farinhas, C. Zerva, M. A. T. Figueiredo, and A. F. T. Martins, "Conformal prediction for natural language processing: A survey," Transactions of the Association for Computational Linguistics, vol. 12, pp. 1497-1516, 2024, doi: 10.1162/tacl_a_00715.

[11] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP," in Proc. 28th Int. Conf. Computational Linguistics, 2020, pp. 757-770, doi: 10.18653/v1/2020.coling-main.66.

[12] B. Wilie et al., "IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding," in Proc. 1st AACL and 10th IJCNLP, 2020, pp. 843-857, doi: 10.18653/v1/2020.aacl-main.85.

[13] A. Conneau et al., "Unsupervised cross-lingual representation learning at scale," in Proc. 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 8440-8451, doi: 10.18653/v1/2020.acl-main.747.

[14] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in Proc. EMNLP-IJCNLP, 2019, pp. 3982-3992, doi: 10.18653/v1/D19-1410.

[15] OASIS Cyber Threat Intelligence Technical Committee, "TAXII Version 2.1," OASIS Standard, Jun. 2021.