Integration of Hybrid Threat Intelligence: The HiPSTer Ontological Method for Cross-Domain Correlation in Influence Operations

Authors and affiliations

Evaldas Bružė[1], R. Andrew Paskauskas[2], Raminta Matulytė[3], Giedrė Sabaliauskaitė[4], Tomas Lavišius[5]

[1] Lithuanian Cybercrime Center of Excellence for Training, Research and Education (L3CE), 08303 Vilnius, Lithuania

[2] Security Research Laboratory, Mykolas Romeris University, 08303 Vilnius, Lithuania

[3] Security Research Laboratory, Mykolas Romeris University, 08303 Vilnius, Lithuania

[4] Faculty of Public Governance and Business, Mykolas Romeris University, 08303 Vilnius, Lithuania

[5] Security Research Laboratory, Mykolas Romeris University, 08303 Vilnius, Lithuania

Abstract

Hybrid threats exploit the interplay between informational, behavioural, psychological, and technical domains, making early detection challenging for security actors. While capabilities in Open-Source Intelligence (OSINT), Social Media Intelligence (SOCMINT), and Natural Language Processing (NLP) have advanced substantially, they remain siloed, leaving meaningful signals fragmented and limiting timely response to Foreign Information Manipulation and Interference (FIMI) campaigns. This paper presents the HiPSTer project's methodological framework addressing this integration gap. The approach frames hybrid threat detection as structured sense-making rather than single-indicator classification, combining empirical use case reconstruction, comparative solution assessment, and ontological modelling. Its core is a semantic integration layer that combines Web Ontology Language (OWL) ontologies, Simple Knowledge Organisation System (SKOS) taxonomies, and Resource Description Framework (RDF) data integration, enabling systematic reconciliation of heterogeneous indicators while accommodating uncertainty.

Assessment of nine existing solutions against 143 functionalities reveals that current platforms, while mature for foundational tasks, lack cross-domain semantic reasoning and dehumanisation detection capabilities that coordinated hybrid operations demand.

The framework is evaluated through two complementary approaches: a case study demonstrating ontological reasoning over the escalation of demonisation in the Russia-Ukraine information environment, supported by query patterns that distinguish coordinated FIMI operations from organic discourse; and multilingual hate speech detection validation across English, Russian, Lithuanian, and Chinese benchmarks. A measurement model combining demonisation indicators with coordination co-indicators provides the logic for graduated threat assessment.

The contribution is methodological. It outlines an approach through which hybrid threat analysis may balance analytical depth with operational usability, support explainability, and align with EU regulatory expectations under the AI Act, GDPR, and Law Enforcement Directive. All ontological artefacts, queries, indicator definitions, and visualisations are available publicly.

Key words

Hybrid threats, Foreign Information Manipulation and Interference (FIMI), open-source intelligence (OSINT), social media intelligence (SOCMINT), ontological framework, dehumanisation detection, multilingual NLP, AI regulatory compliance

Introduction

Hybrid threats have become increasingly difficult to isolate, not because individual indicators are invisible, but because they tend to surface across domains that are traditionally analysed separately. Narrative shifts coordinated online behaviour, and subtle changes in information ecosystems often precede more overt actions, yet they rarely trigger attention when viewed in isolation. In practice, this creates a persistent analytical gap for law enforcement and security practitioners: signals are present, but they do not coalesce into actionable insights early enough.

Much of the existing research on hybrid threats focuses either on strategic-level doctrines or on narrowly defined technical problems, such as disinformation detection or bot identification. While these contributions are valuable, they often assume relatively stable threat patterns or well-defined adversarial signatures. What we encountered early in the HiPSTer project was a different reality. Signals were fragmented, context-dependent, and frequently ambiguous. Even when advanced analytical tools were available, practitioners struggled to relate outputs to operational decision-making in a way that felt timely or proportionate.

Recent studies have shown that hybrid and information operations increasingly rely on adaptive, low-visibility techniques that exploit social, cognitive, and informational vulnerabilities (Bizouarn, 2023) rather than direct technical compromise (Bennett & Livingston, 2018; Starbird et al., 2020). From an analytical perspective, this shifts the problem from detecting discrete events to interpreting evolving patterns. The challenge is not only technical, but methodological: how to structure analysis so that weak signals can be assessed without amplifying noise or bias.

Within this context, the HiPSTer project set out to explore whether a more integrated analytical approach could support earlier and more reliable identification of hybrid threat dynamics. Rather than treating OSINT, behavioural indicators, and contextual knowledge as separate analytical layers, the project investigates how these elements can be reasoned over jointly. This integration requires treating uncertainty as a first-class consideration, accommodating partial confidence and evolving interpretations rather than forcing binary classifications.

One issue we faced early on was the tension between analytical depth and operational usability. Highly detailed models can capture nuance, but they often become opaque to practitioners who need to make decisions under time pressure. Conversely, overly simplified indicators risk missing the very dynamics that make hybrid threats effective. This paper reflects that tension directly. The approach described here does not claim to eliminate ambiguity; rather, it aims to make ambiguity visible and manageable within an operational workflow.

The contribution of this paper is therefore methodological rather than purely technical. We describe how the HiPSTer approach frames hybrid threat detection as a process of structured sense-making, grounded in real-world constraints faced by law enforcement and security actors. By situating analytical choices within both existing research and practical use cases, we aim to clarify what such systems can reasonably support – and where their limits remain.

The paper begins by outlining the methodological approach guiding the research. From there, it provides essential background, reviewing the state of the art in open-source intelligence (OSINT), social media intelligence (SOCMINT), and natural language processing (NLP) for hybrid threat detection, examining the adversarial context established by Russian and Chinese information operations, assessing existing solutions against operational requirements, and establishing the regulatory landscape – including the General Data Protection Regulation (GDPR), the AI Act, and the Law Enforcement Directive (LED) – that constrains system design.

A key focus is placed on the system architecture and ontological framework, detailing how the HiPSTer knowledge graph enables cross-domain semantic reasoning through a three-layer architecture combining Web Ontology Language (OWL) ontologies, Simple Knowledge Organization System (SKOS) taxonomies, and Resource Description Framework (RDF) data integration. The case study on demonization escalation in the Russia-Ukraine information environment demonstrates the framework's capabilities, illustrating how it captures coordinated hybrid threat patterns through SPARQL Protocol and RDF Query Language (SPARQL) queries that distinguish orchestrated Foreign Information Manipulation and Interference (FIMI) operations from organic discourse. Empirical validation of multilingual hate speech detection across English, Russian, Lithuanian, and Chinese benchmarks – achieving Area Under the Receiver Operating Characteristic Curve (AUC-ROC) scores of 0.67–0.93 – provides quantitative evidence that linguistic indicators populating the ontology's Dehumanization and Demonization domain can be reliably detected across diverse linguistic contexts.

The concluding discussion addresses limitations and future directions, emphasising the pathway from Technology Readiness Level 4 (TRL-4) validation to operational deployment, the potential to extend multimodal integration, and the need for broader empirical validation across additional geopolitical contexts.

Significantly, this paper presents a methodological and architectural research contribution rather than a conventional hypothesis–experiment–result study. The HIPSTer framework does not aim to produce statistically generalisable performance claims or ground-truth attribution. Instead, it formalises a structured sense-making approach capable of integrating heterogeneous indicators in the face of the uncertainty that characterises real-world hybrid threat environments. The contribution lies in demonstrating that ontological reasoning can bridge the analytical silos identified in current OSINT, SOCMINT, and NLP research, enabling cross-domain correlation that classification-only approaches cannot. Validation is achieved through a detailed case study that applies the framework to documented influence operations in the Russia-Ukraine information environment, assessing whether the methodology supports more interpretable, structured analytical reasoning than current state-of-the-art tools.

Methodological approach

The methodology presented in this paper constitutes its primary contribution. The following sections define the analytical logic and ontological principles of the HIPSTer framework, independent of specific implementation choices. The goal is not to produce statistically generalisable performance metrics, but to formalise a structured approach for integrating heterogeneous threat indicators under conditions of uncertainty. Specific tools, platforms, and technical configurations described in subsequent sections serve as illustrative implementations and do not constrain the methodology's general applicability.

The methodological approach developed within the HiPSTer project is grounded in the observation that hybrid threats rarely manifest as isolated technical events. Instead, they emerge through the gradual alignment of informational, behavioural, psychological, and societal signals that only become meaningful when interpreted in relation to one another. Existing analytical practices tend to fragment these signals across disciplinary and technological boundaries, which significantly limits early detection and coherent risk assessment. The methodology described here was designed explicitly to address this structural limitation.

Our starting point was empirical rather than theoretical. Before formalising any analytical model, the project undertook a systematic reconstruction of real-world hybrid threat cases, focusing primarily on Russian and Chinese influence and information operations. These historical analyses revealed recurring patterns: weak signals appeared early, often dispersed across platforms and domains, while overt indicators surfaced only later, once campaigns had already gained momentum. This finding shaped a core design principle of the methodology: analytical processes must prioritise signal reconciliation rather than single-indicator detection.

From this empirical foundation, the methodology evolved into a multi-stage analytical workflow that integrates use case analysis, state-of-the-art review, indicator construction, solution assessment, and ontological modelling into a single coherent system. Each stage informs the others, and none is treated as an independent or purely preparatory activity. This integrated structure reflects a deliberate departure from linear threat-analysis pipelines, which tend to hard-code assumptions early and struggle to adapt to evolving threat dynamics.

A central methodological challenge concerned the balance between analytical depth and operational usability. Rather than resolving this tension by choosing one over the other, the methodology accommodates both by structuring indicators hierarchically across macro, meso, and micro levels. Macro-level indicators capture strategic patterns and campaign-level dynamics; meso-level indicators reflect behavioural and narrative alignment; micro-level indicators focus on specific technical, linguistic, or activity-based signals. This layered approach enables analysts to move between abstraction levels without losing contextual coherence.

Another key methodological decision was to treat uncertainty as an explicit analytical variable rather than an error to be eliminated. Hybrid threat analysis frequently operates under conditions of partial information, deliberate deception, and ambiguous intent. The methodology accordingly aggregates weighted indicators and contextual factors into evolving risk assessments that can be updated as new information becomes available.

Semantic integration plays a critical role in enabling this process. The methodology introduces a formal hybrid threat ontology that encodes actors, behaviours, narratives, indicators, and contextual relationships within a unified conceptual structure. This ontological layer is not an abstract modelling exercise; it functions as the backbone for automated reasoning, explainability, and cross-domain correlation. By formalising relationships that are often handled implicitly by human analysts, the ontology enables systematic reconciliation of heterogeneous indicators drawn from OSINT, SOCMINT, NLP outputs, and cyber intelligence sources.

The integration of existing technological solutions was approached cautiously. Rather than assuming that state-of-the-art tools could simply be combined, the project conducted a structured assessment of commercial, institutional, and research-driven systems to identify functional gaps and methodological blind spots. This analysis confirmed that while many tools perform well within narrow domains – such as metadata extraction, sentiment analysis, or network mapping – few support cross-domain semantic reasoning or hybrid-threat-specific interpretation. These findings directly informed the design boundaries of the HiPSTer methodology and are detailed in the background section that follows.

The methodology was designed with operational deployment in mind from the outset. Analytical components were evaluated not only for conceptual soundness but also for scalability, explainability, and compatibility with regulatory constraints, including emerging European requirements under the AI Act and GDPR. This operational orientation influenced decisions such as modular architecture, human-in-the-loop integration, and transparency of analytical outputs. As a result, the methodology supports deployment at TRL 7–8 levels without requiring extensive post hoc adaptation. The regulatory context shaping these design choices is examined in the regulatory perspective section.

Empirical validation constitutes a further methodological priority. Beyond the qualitative case study demonstrating ontological reasoning over demonisation escalation patterns, the framework incorporates quantitative validation of linguistic indicator detection. Multilingual hate speech classification – validated across English, Russian, Lithuanian, and Chinese benchmarks – provides empirical evidence that the indicators populating the Dehumanization and Demonization domain (SKOS Domain 11) can be reliably detected across diverse linguistic contexts. This validation addresses the requirement for reproducible, machine-assisted analysis central to the HiPSTer approach.

Taken together, the HiPSTer methodological approach represents a shift from tool-centric analysis toward structured sense-making. Rather than asking whether a specific signal indicates malicious activity, the methodology asks how multiple signals align over time, across domains, and within known hybrid threat patterns. This reframing is essential for addressing the adaptive and low-visibility nature of contemporary hybrid operations and forms the conceptual foundation for the system architecture and analytical workflows described in the following sections.

State of the art and beyond in hybrid threat intelligence

This section reviews the current state of research and practice across the three analytical domains most relevant to hybrid threat detection: open-source intelligence (OSINT), social media intelligence (SOCMINT), and natural language processing (NLP). For each domain, we examine recent advances, identify persistent limitations, and explain how the HiPSTer ontological framework addresses integration gaps left unresolved by current approaches. The section concludes by examining the adversarial context – specifically Russian and Chinese information operations – that establishes the operational requirements against which any defensive framework must be assessed.

Open-source intelligence (OSINT)

A comprehensive review of OSINT capabilities and integration challenges for hybrid threat detection has been presented in Paskauskas et al. (2026), forthcoming in the Journal of Intelligent Communication (JIC). That review provides a detailed analysis of contemporary OSINT research across nine academic studies, examining capabilities in explainable AI integration for botnet detection (Suryotrisongko et al., 2022a, 2022b), LLM-based tool orchestration for autonomous threat intelligence harvesting (Yuan et al., 2024), real-time feed integration converting 49 public data streams into intrusion detection rules (Vacas et al., 2018), and multimodal forensics including steganographic malware detection (Fauziyyah et al., 2023). The review extensively evaluates commercial OSINT platforms including Babel Street's multilingual monitoring across 200+ languages and 30+ social media platforms, WebIQ's Voyager Suite (2025) with Visual Media Analytics for semantic image

search and darknet crawling, enterprise solutions (Palantir Gotham[1], IBM i2 Analyst's Notebook[2]), and open-source tools (Maltego, Python libraries including Scrapy, Pandas, Matplotlib). Significantly, the analysis identifies three fundamental limitations when addressing hybrid threat scenarios: domain-specific analytical silos, limited contextual integration, and the absence of multi-modal reasoning frameworks. These gaps reveal that while individual tools demonstrate sophisticated capabilities within bounded domains, no current OSINT framework provides the systematic cross-domain correlation capabilities necessary to detect coordinated hybrid operations.

Since that analysis, significant developments have emerged that both extend and complicate the OSINT landscape for hybrid threat detection. The integration of generative AI and large language models into OSINT workflows represents a fundamental shift in capabilities and challenges. Avrahami, Zwilling, and Hajaj provide a strategic framing of OSINT's evolving role in proactive cybersecurity, explicitly addressing integration challenges within organisational intelligence frameworks while balancing cost-effectiveness against persistent concerns about data reliability and legal compliance (Avrahami et al., 2025), issues particularly salient for operational deployment of hybrid threat detection systems.

However, this technological advancement confronts a critical challenge: the proliferation of AI-generated synthetic media and deepfakes fundamentally undermines the foundational OSINT principle of "trust but verify." Recent research documents a 550% growth in deepfake videos since 2019, with projections reaching 8 million by 2025, while ENISA (2025) reports that AI-generated content now accounts for over 80% of observed social engineering attacks. Detection capabilities lag behind generation sophistication, with recent studies reporting detection accuracy of only 0.73 and precision of 0.75 for synthetic media identification, contributing to projected global damages of $40 billion by 2027 from deepfake-enabled fraud and manipulation (Balogun et al., 2025).

Beyond technical detection challenges, recent work highlights emerging requirements for OSINT practitioners. Social Links argues that effective OSINT in 2025 requires sophisticated provenance analysis (Social Links, 2025), using digital forensics and cross-platform signal correlation rather than accepting media at face value, particularly as users migrate from centralised platforms to fragmented ecosystems (Telegram, Discord, Mastodon) that complicate systematic monitoring (Yadav et al., 2023). Industry analyses reinforce this multi-dimensional challenge: commercial platforms report that AI-driven OSINT must now process millions of data points in real-time across dark web forums, social media, and news sources while simultaneously validating authenticity, tracking coordinated campaigns, and maintaining compliance with evolving regulatory frameworks, including the EU AI Act and data protection requirements (Fivecast, 2025; Authentic8, 2025).

These developments reinforce rather than resolve the three critical gaps identified in the JIC comprehensive review: domain-specific analytical silos now extend to AI-generated content detection; limited contextual integration becomes more acute as synthetic media blurs the boundaries between authentic and manipulated information across platforms; and the absence of multi-modal reasoning frameworks persists despite advances in individual detection capabilities. The fundamental challenge remains unchanged but intensified: hybrid threats exploit the seams between analytical disciplines, and current OSINT approaches – despite impressive AI integration – continue to **analyse** these domains in isolation rather than providing systematic cross-domain correlation for coordinated campaign detection.

HiPSTer's ontological integration approach

HiPSTer's ontology-driven approach addresses these limitations by integrating semantics rather than simply aggregating existing analytical capabilities. Where current OSINT research optimises individual detection components, HiPSTer's framework enables systematic correlation of threat indicators across the multiple domains that hybrid threats characteristically exploit. The ontological foundation provides formal knowledge representation for reasoning about relationships between seemingly disparate indicators, such as correlating domain generation algorithm (DGA) activity patterns with coordinated social media campaigns or linking infrastructure vulnerabilities to broader influence operations (Suryotrisongko et al., 2022a, 2022b).

This semantic integration approach represents a conceptual advancement beyond current OSINT literature: rather than developing another specialised analytical tool, the framework orchestrates diverse intelligence sources and analytical methods within a unified reasoning environment specifically designed for the multi-domain, coordinated nature of contemporary hybrid threat campaigns. The ontology enables persistent modelling of threat actor tactics, techniques, and procedures across domains, supporting attribution and campaign tracking capabilities that exceed the scope of current domain-specific OSINT research.

Social media intelligence (SOCMINT)

A comprehensive review of SOCMINT capabilities for hybrid threat detection has been presented in JIC (Paskauskas et al., 2026). That review provides detailed analysis distinguishing between studies explicitly addressing hybrid threats as coordinated, multi-domain security phenomena, including critical assessment showing Western powers losing the information component of hybrid conflicts (Dover, 2020); how a framework can detect radical behavioural traits (Cardenas et al., 2019); as well as instruments for countering hybrid security threats (Mothe et al., 2020) and research employing advanced methodologies for traditional cybersecurity problems such as hybrid deep learning architecture for bot detection achieving impressive performance metrics (Ellaky et al., 2024); or LSTM and BERT models for darknet forum analysis (Sangher et al., 2024); not to mention the MARPLE framework for terrorist threat intelligence (Biagio et al., 2024) and the

[1] https://www.palantir.com/platforms/gotham
[2] https://www.ibm.com/support/pages/download-ibm-i2-analysts-notebook-931?utm_source=copilot.com

comparative analysis of various AI approaches (Dragos et al., 2020, 2022). The JIC review extensively evaluates commercial SOCMINT platforms, including Babel Street's multilingual monitoring across 200+ languages and 30+ social media platforms, WebIQ's Voyager Suite extending capabilities to encrypted messaging and darknet forums with Visual Media Analytics, and specialised platforms focused on coordinated inauthentic behaviour, including Graphika's network[3] mapping of state-linked influence operations (Russian Internet Research Agency (IRA), Chinese state-aligned networks, Iranian campaigns) and Botometer's research-standard bot-likelihood scoring (Ellaky et al., 2024). Significantly, the analysis identifies three fundamental limitations: conceptual confusion between hybrid threats and hybrid methodologies, analytical fragmentation across threat domains, and absence of cross-platform and cross-domain intelligence integration. Dover's critique reinforces this, noting that while SOCMINT provides horizon-scanning capabilities, it offers insufficient real-time warning capacity during rapidly evolving hybrid conflicts, precisely when coordinated threat correlation becomes most critical.

Since the JIC analysis was undertaken, significant developments have emerged that both extend and challenge the SOCMINT landscape for hybrid threat detection. The most substantial advancement concerns cross-platform coordination detection, addressing a critical blind spot in traditional single-platform analysis. Empirical evidence has been put forward (Luceri et al., 2025) of coordinated inauthentic activity spanning X (formerly Twitter), Facebook, and Telegram during the 2024 U.S. Presidential Election, revealing systematic Russian-affiliated media promotion across platforms and demonstrating that coordination frequently transcends platform boundaries, precisely the cross-platform integration gap identified in the JIC comprehensive review. The advanced coordination detection model uncovers substantial intra- and cross-platform coordinated activity driving highly partisan, low-credibility, and conspiratorial content, with QAnon-related narratives particularly prevalent on Telegram. Significantly, coordinated actors on Telegram relied substantially more on AI-generated content than organic users did, while on Facebook, the opposite pattern held, suggesting platform-specific manipulation strategies that complicate unified detection frameworks.

However, questions about SOCMINT's actual influence capabilities have emerged alongside these technical advances. A challenge to fundamental assumptions about the effectiveness of coordinated inauthentic behaviour was issued, finding that despite high coordination levels in the German COVID-19 #nocovid discourse, coordinated tweets did not trigger increased engagement from non-coordinated users (Sarhan & Hegelich, 2025). Granger causality analysis failed to establish causal links between coordinated activity and subsequent rises in non-coordinated discourse, suggesting that coordination may primarily serve as an amplification mechanism within pre-existing ideological groups rather than expanding reach to new audiences, raising critical questions about whether current SOCMINT detection priorities align with actual threat impact. This finding reinforces Dover's earlier critique that SOCMINT offers limited real-time warning during rapidly evolving conflicts, now extending the concern to whether the detection of coordination itself meaningfully indicates influence success.

Bot-detection capabilities have evolved significantly through the integration of graph neural networks and semi-supervised learning. Liu et al. (2025) trace the evolution from traditional machine learning techniques through deep learning architectures to contemporary group-level detection paradigms, with emerging Large Language Model (LLM)-based detection representing a new frontier. Others have demonstrated how semi-supervised approaches using Relational Graph Attention Transformers that incorporate characteristics of the social environment address the persistent challenge that coordination and automation remain orthogonal concepts (D. Huang et al., 2025), revealing that coordinated accounts are not necessarily automated, and automated accounts may not exhibit coordination patterns. However, these advances continue to operate within platform-specific analytical frameworks, perpetuating the domain fragmentation identified in the original review rather than enabling the cross-platform reasoning necessary for hybrid threat detection.

Rogers and Righetti (2025) provide crucial conceptual clarification by offering a typology of coordinated link-sharing networks on Facebook, identifying six distinct types: legitimate media groups, advertising networks, large public groups, political support/opposition networks, and genuine influence operations. This taxonomy reveals that "coordinated behaviour" encompasses diverse phenomena – from benign community organising to malicious manipulation – challenging the assumption that coordination detection alone identifies inauthentic or harmful activity. The typology exposes analytical imprecision in conflating coordination with inauthenticity, precisely the "conceptual confusion between hybrid threats and hybrid methodologies" identified as a fundamental limitation in the comprehensive review.

These developments reinforce rather than resolve the three critical gaps identified in the JIC comprehensive review: conceptual confusion between coordination as methodology versus threat indicator now extends to distinguishing between legitimate and malicious coordination; analytical fragmentation across platforms intensifies as actors develop platform-specific strategies (different AI-generated content usage on Telegram versus Facebook); and the absence of cross-platform intelligence integration persists despite evidence that threat actors systematically exploit platform boundaries. The fundamental challenge identified by Dover remains: current SOCMINT approaches provide valuable horizon-scanning within bounded contexts but cannot automatically correlate social media signals across platforms or with the cyber-technical indicators necessary to detect coordinated hybrid operations. Sarhan and Hegelich's finding that detected coordination may not meaningfully influence broader discourse adds a troubling dimension – current detection capabilities may identify coordination that matters little while missing influence mechanisms that matter greatly.

[3] https://www.graphika.com/how-it-works

HiPSTer's ontological integration approach

HiPSTer's framework addresses these gaps through semantic reasoning rather than advancing individual SOCMINT detection capabilities. Where current research optimises specialised analytics within social media domains, the ontological approach enables systematic correlation between social media intelligence and other intelligence sources within unified knowledge representations designed specifically for multi-domain threat reasoning.

The framework is designed to integrate the sophisticated detection capabilities demonstrated across current SOCMINT research – from bot detection algorithms (Ellaky et al., 2024) to extremist content analysis (Biagio et al., 2024) to darknet monitoring (Sangher et al., 2024) – within a semantic environment that can reason about relationships between social media indicators and broader hybrid threat campaigns. This represents a conceptual shift from optimising individual social media analytics to orchestrating diverse intelligence capabilities within a coherent system capable of detecting coordination patterns spanning information, cyber, and physical domains.

Natural language processing (NLP)

A comprehensive review of NLP capabilities for hybrid threat detection has been presented in the JIC review article discussed previously (Paskauskas et al., 2026). That review provides a detailed analysis of contemporary NLP research, including methodological approaches and performance metrics across eleven studies spanning knowledge-based entity extraction with KnowCTI (Wang et al., 2024), automated report generation with AGIR (Silver et al., 2022), multimodal fact-checking with SNIFFER (Qi et al., 2024), and agentic verification systems utilising FactAgent (Li et al., 2024). The analysis examines in depth the implications of generative AI and deepfakes for hybrid threat operations, noting that AI-generated content now accounts for over 80% of observed social engineering attacks (ENISA, 2025).

Significantly, the JIC review contextualises these technical capabilities within European regulatory frameworks, including the General-Purpose AI Code of Practice (European Commission, 2025b) and the Commission's Guidelines on Prohibited AI Practices (European Commission, 2025a), demonstrating how governance constraints shape both adversarial content generation and defensive detection systems. The analysis identifies three critical gaps: absence of cross-domain reasoning frameworks, limited operational explainability and attribution, and fragmentation of detection capabilities across threat components.

Since that analysis, significant developments have emerged that simultaneously demonstrate advances in specialised NLP capabilities while reinforcing the fundamental integration challenges identified in the comprehensive review. The integration of LLMs into threat detection workflows represents the most substantial shift, though with mixed results that illuminate persistent limitations in operational deployment.

Performance evaluations of LLMs for propaganda detection reveal critical gaps between generalist capabilities and specialised threat analysis requirements. Investigators conducted systematic evaluations of GPT-3.5, GPT-4, and Claude 3 Opus on detecting six common propaganda techniques in news articles, finding that while GPT-4 demonstrated superior performance (F1=0.16) compared to other LLMs, it failed to outperform a RoBERTa-CRF baseline (F1=0.67) by substantial margins (Jose & Greenstadt, 2024). The models performed relatively well on common techniques like name-calling and loaded language, but showed declining accuracy with more complex tactics such as appeal to fear and flag-waving, precisely the sophisticated manipulation tactics characteristic of state-sponsored hybrid operations. More promising results through strategic prompt engineering for Arabic propaganda span annotation were achieved, demonstrating that providing annotation context improves GPT-4 performance compared to human annotators and achieves state-of-the-art results when serving as an expert consolidator (Hasanain et al., 2024). However, this success required language-specific optimisation and carefully curated training data, reinforcing rather than resolving the resource intensity limitations identified in the comprehensive review.

Conversely, research examining LLMs' capacity to generate propaganda reveals disturbing asymmetries between defensive detection capabilities and offensive generation potential. Ten LLMs across 20 disinformation narratives were evaluated, concluding that current models – including open-source variants – can generate convincing news articles aligned with dangerous disinformation narratives, complete with real or hallucinated supporting evidence (Goldstein et al., 2024). Researchers traced rhetorical strategies across GPT-3.5-Turbo, GPT-4o, and GPT-4.1, generating over 340,000 articles on politically divisive topics and achieving Area Under the Receiver Operating Characteristic curve (AUROC) above 98% in distinguishing model-generated propaganda from human-written content (Shulman et al., 2025). All three models rely heavily on cognitive language to simulate deliberation, while maintaining consistent moral framing, though each version employs distinct stylistic choices. This generation-detection asymmetry creates an operational disadvantage: adversaries can rapidly generate sophisticated propaganda at scale while defenders struggle to detect it reliably, particularly when generation techniques evolve across model versions.

Advances in fact-checking automation demonstrate incremental progress while highlighting persistent challenges in explainability and integration. Li et al. (2024) integrated Retrieval-Augmented Generation (RAG) systems with LLMs for COVID-19 fact-checking, constructing a context dataset of approximately 130,000 peer-reviewed papers from PubMed and Scopus. The RAG-enhanced approach substantially improved accuracy and contextual relevance compared to standalone LLMs by reducing hallucinations, providing source citations for transparency, and addressing the explainability gap identified in the comprehensive JIC review. However, this approach requires domain-specific knowledge bases and remains constrained to topics with substantial peer-reviewed literature, limiting applicability to emerging hybrid threats lacking established documentation. Hamzic et al. (2025) developed a unified pipeline combining named entity recognition (NER), relationship extraction, classification, and summarisation for Cyber Threat Intelligence (CTI) and OSINT analysis. Critically, their evaluation revealed that general-purpose tokenizers recognise only 1.62% of specialised MITRE ATT&CK terms –

quantifying the domain-specific vocabulary gap that undermines generalist NLP approaches for threat intelligence. Their pipeline integrates symbolic knowledge graphs with semantic vector search, storing text segments in vector databases enriched with metadata while maintaining relation mappings in graph databases, enabling both semantic retrieval and structured reasoning. The approach surpasses state-of-the-art models, including TRAM and TTPXHunter, for Tactics, Techniques, and Procedures (TTP) extraction, demonstrating that domain-specific semantic frameworks outperform general-purpose models for specialised threat analysis. However, the pipeline remains focused on cyber threat intelligence within technical domains, rather than extending to the cross-domain correlation needed to detect coordinated campaigns spanning information operations, cyber activities, and physical threats that characterise hybrid operations.

These developments reinforce rather than resolve the three critical gaps identified in the comprehensive review. The absence of cross-domain reasoning frameworks intensifies as research demonstrates strong performance within bounded domains (Arabic propaganda, COVID-19 fact-checking, CTI analysis) but limited transferability across threat contexts – Hamzic et al.'s finding that general tokenizers recognise only 1.62% of domain terms suggests this specialisation challenge extends across threat intelligence disciplines. Limited operational explainability persists despite RAG-enhanced transparency; fragmentation of detection capabilities across threat components continues as specialised models optimise individual tasks (propaganda detection, fact-checking, TTP extraction), while systematic integration frameworks remain absent. No current approach provides a unified analytical environment capable of correlating propaganda narratives with cyber infrastructure exploitation and physical threat indicators simultaneously.

The fundamental challenge articulated across this literature aligns precisely with the limitation identified in the JIC comprehensive review: hybrid threats manifest through coordinated operations across multiple domains simultaneously, yet current NLP research advances individual analytical components in isolation. Hamzic et al.'s architectural integration of symbolic reasoning with semantic search within the CTI domain demonstrates the viability of unified frameworks, but extending this integration across information operations, cyber threat intelligence, and physical security domains – the operational requirement for hybrid threat detection – remains an unsolved research challenge that HiPSTer's ontology-driven approach specifically targets.

Addressing NLP Integration Gaps Through Ontological Frameworks

While existing research demonstrates impressive capabilities within specific domains such as entity extraction (KnowCTI), automated reporting (AGIR), multimodal fact-checking (SNIFFER), and vulnerability assessment (Marchiori et al., 2023), these advances remain fundamentally siloed. Current approaches require analysts to manually integrate outputs from disparate tools, creating bottlenecks in threat detection and limiting the ability to identify coordinated campaigns that span multiple domains. The HiPSTer ontology-driven framework is designed to address four critical integration challenges identified across the literature:

Semantic Integration of Multi-Modal Intelligence: Unlike existing point solutions, HiPSTer's ontological framework enables systematic correlation between heterogeneous data sources. Where current systems process dark web intelligence (Pasupuleti, 2022) or visual misinformation (Qi et al., 2024) in isolation, the ontology-based approach allows threat indicators from text, images, network logs, and social graphs to be semantically linked within a unified knowledge representation. This addresses the fundamental limitation that hybrid threats manifest across multiple information channels simultaneously.

Automated Cross-Domain Reasoning: Building on agent-based approaches like FactAgent, HiPSTer's framework orchestrates specialised detection modules through ontology-guided workflows. Rather than requiring manual coordination between tools like KnowCTI for entity extraction and CVE scorers for vulnerability assessment, the system's semantic reasoning engine automatically determines appropriate analytical pathways based on threat context and evidence patterns, addressing the workflow integration gap highlighted by Huang et al.

Real-Time Hybrid Threat Correlation: The framework's design specifically targets the cross-domain detection blind spot evident throughout the literature. By maintaining persistent semantic models of threat campaigns, HiPSTer can automatically identify when seemingly unrelated events, such as coordinated disinformation campaigns and emerging cyber vulnerabilities represent components of broader hybrid operations. This capability directly addresses the limitation that current state-of-the-art approaches cannot systematically connect threats across information, cyber, and physical domains.

Operationally-Ready Transparency and Governance: Recognising the explainability challenges identified across multiple studies, HiPSTer's ontological approach inherently supports provenance tracking and semantic explanation generation. Unlike post-hoc interpretability methods, the framework's knowledge-based reasoning provides natural explanation pathways while enabling policy-compliant information handling – a critical requirement for operational deployment that remains under-addressed in current research.

The preceding review reveals a consistent pattern across OSINT, SOCMINT, and NLP domains: while individual analytical capabilities have achieved considerable sophistication, hybrid threats systematically exploit the seams between these disciplines. Current tools excel within bounded specialisations but lack the cross-domain semantic reasoning necessary to detect coordinated campaigns spanning information, cyber, and physical domains. Before detailing how the HiPSTer framework addresses these integration gaps at the architectural level, it is necessary to examine the adversarial context that establishes operational requirements – specifically, the documented tradecraft of Russian and Chinese state-sponsored information operations that any defensive system must be able to counter.

Processing adversarial context and operational requirements

Effective hybrid threat detection requires understanding the operational tradecraft that defensive systems must counter. Russian and Chinese state-sponsored information operations represent the most sophisticated and persistent threats facing European security, establishing the benchmark against which detection capabilities must be assessed. Analysis of documented campaigns reveals recurring patterns that inform defensive framework design.

Russian information warfare doctrine integrates military, intelligence, and psychological operations into a unified concept of "information confrontation," wherein narrative control constitutes a domain of conflict coequal with kinetic operations (Borisov, 2024). Operations in Ukraine since 2014 demonstrate this integration: military actions coordinate with cyber-attacks against critical infrastructure, targeted disinformation campaigns on social platforms, and strategic communications through state-controlled media to shape international perceptions while degrading adversary effectiveness (Kalensky & Osadchuk, 2024). VKontakte and Telegram serve as primary platforms for both OSINT collection and influence operations, exploiting widespread adoption in post-Soviet spaces and permissive content moderation policies. The IRA's 2016 operations targeting U.S. elections established templates subsequently refined: coordinated inauthentic behaviour employing bot networks to amplify divisive content, authentic-appearing personas building credibility before deploying disinformation, and cross-platform coordination ensuring narrative saturation before fact-checking responses emerge (Eady et al., 2023). The Portal Kombat network identified by France's VIGINUM service in 2024 exemplifies continued evolution: over 1,000 websites across European languages promoted coordinated pro-Russian narratives through culturally tailored content exploiting specific national debates (VIGINUM, 2024, 2025naas).

Chinese military information operations reflect distinct strategic priorities while increasingly adopting techniques pioneered in Russian operations. The People's Liberation Army conceptualises information warfare through the "Three Warfares" framework: psychological warfare targeting adversary decision-maker perceptions, media warfare shaping domestic and international narratives, and legal warfare establishing favourable interpretations of international norms (Cochran, 2020). Rather than employing bot networks extensively, Chinese operations often leverage genuine accounts – diaspora communities, nationalist volunteers, and coordinated authentic behaviour by users responding to official guidance (Charon & Jeangène Vilmer, 2021). TikTok's algorithm-driven content distribution enables influence operations that exploit platform mechanisms rather than requiring overt coordination: strategically placed content that resonates with target audiences receives organic amplification, effectively weaponising commercial recommendation systems (Finkelstein et al., 2025).

The European Union's institutional framing explicitly identifies FIMI as a primary hybrid threat vector, characterised by coordinated campaigns employing inauthentic news articles, fabricated investigations, AI-generated deepfakes, and cross-platform narrative manipulation designed to interfere in elections, discredit public institutions, and erode democratic resilience (EEAS, 2026). ENISA's 2025 Threat Landscape documents that AI-generated content now accounts for over 80% of observed social engineering attacks, while approximately a quarter of FIMI content focuses specifically on degrading the Union through negative narratives targeting France, Germany, and Poland (ENISA, 2025).

This adversarial landscape establishes specific operational demands for defensive systems. The comprehensive capability assessment in Paskauskas et al. (2026) identifies persistent gaps across OSINT, SOCMINT, and NLP domains: while individual analytical disciplines achieve high technical sophistication within bounded specialisations, current defensive systems remain siloed, lacking the cross-domain semantic reasoning necessary to correlate technical cyber indicators with coordinated narrative manipulation, distinguish genuine hybrid threats from hybrid analytical methodologies, and support attribution under conditions of deliberate ambiguity. These architectural limitations – rather than deficits in individual detection capabilities – motivate the ontological integration approach that HiPSTer's framework addresses.

Comparative assessment of existing solutions

To systematically evaluate the landscape of hybrid threat detection capabilities, the HiPSTer project conducted a structured assessment of nine solutions against operational requirements derived from documented threat patterns. The methodology employed expert-based evaluation, with each solution rated by two independent assessors on a four-point scale (0 = absent, 1 = initial capabilities, 2 = present, 3 = well developed) across 143 detailed functionalities grouped into 12 high-level capability domains.

The solutions assessed encompassed commercial platforms (Babel Street, WebIQ, Logically, Storyzy), EU-level initiatives (EEAS FIMI activities, EU-NATO Fusion Cell), national capabilities (France's VIGINUM), Lithuania's (NAAS, 2023), and research projects (STARLIGHT, 2025). Materials gathered included provider documentation, independent reviews, and operational evaluations. The results are summarised in Table 1.

Table 1. [was removed from here]

While Table 1 provides a solution-by-solution overview, the critical question for hybrid threat detection is whether the current landscape adequately covers the analytical capabilities that coordinated influence operations demand. To address this, the assessment aggregated functionality ratings across all twelve capability domains, revealing where collective maturity exists and where systemic gaps persist regardless of individual platform strengths. Table 2 presents this capability domain coverage across all assessed solutions.

Table 2. [was removed from here]

The assessment reveals a consistent pattern. Solutions demonstrate mature capabilities for foundational functions – metadata extraction, linguistic analysis, narrative identification, and influencer network mapping. However, significant gaps persist in precisely the capabilities that hybrid threat detection demands: cross-platform behavioural correlation, semantic reasoning beyond keyword matching, dehumanization and demonization detection, and psychological pattern analysis.

The Dehumanization and Demonization domain represents a particularly critical gap. As documented in the case study that follows, escalating demonization rhetoric frequently precedes and accompanies hybrid threat campaigns, serving as both an indicator of coordinated operations and a mechanism for mobilising hostile action. Yet current solutions offer minimal automated capability to detect such content, relying instead on keyword-based approaches that miss contextual and coded expressions of dehumanising discourse.

These findings directly informed HiPSTer's design priorities: the ontological framework specifically addresses the semantic reasoning and cross-domain correlation gaps that incremental improvements to existing platforms cannot resolve. The SKOS taxonomy incorporates detailed functionality classifications across all twelve domains, with particular emphasis on the underserved Dehumanization and Demonization category that current commercial and institutional solutions fail to address adequately. The complete detailed functionality assessments underlying this summary are available in the project's GitHub repository (https://github.com/SecOntologyLab/hipster-ontology).

Regulatory perspective

While systems dedicated to hybrid threat analysis prioritise analytical depth with operational usability, deployment requires compliance with regulatory expectations for AI-supported security systems. For HiPSTer-like systems developed in the EU, applicable legislation spans multiple dimensions, for example, jurisdiction (national, European Union, international), purpose (civilian, national security/defence, law enforcement, or mixed), and binding force (hard law versus soft law/policy). This creates a complex regulatory mapping that system designers must navigate from the outset.

The following provides a non-exhaustive overview of the regulatory landscape influencing the development of HiPSTer-like systems. Where relevant, Lithuanian national laws are referenced as illustrative case examples.

Protection of fundamental rights

HIPSTer-like systems, while beneficial for further development of efficient tools to combat hybrid threats, may pose significant threats to fundamental rights. Such rights are primarily established in the European Convention on Human Rights[4], the Charter of Fundamental Rights of the European Union, and national constitutions. For example, considering the amount of personal data to be processed as part of the HIPSTer operational necessity, the right to respect for private and family life or the right to data protection can be affected. Furthermore, given the vast amount of content published and processed through SOCINT and the related risks of false flags, freedom of expression may also be at risk, and the deployment of HIPSTer-like systems may cause a 'chilling effect'.[5] Considering the profiling aspect of HIPSTer-like systems, the prohibition of discrimination may be affected. Therefore, not only must regulatory requirements be closely followed, but, in developing the technological capacities of HIPSTer-like systems, fundamental rights must be considered at every step of system design and implementation.

Data protection frameworks

As HiPSTer-like systems process substantial volumes of personal data, the EU General Data Protection Regulation (GDPR) 2016/679[6] constitutes the foundational regulatory framework. The principle of privacy-by-design (Art. 25) requires that appropriate technical and organisational measures implementing data-protection principles be established at the time processing means are determined – a requirement that informed the HiPSTer architecture's design from inception. HiPSTer-like systems engage in multiple forms of personal data processing – OSINT collection, AI-based profiling, partner data sharing, and ML training dataset creation – each triggering specific GDPR obligations regarding lawful basis, purpose limitation, data minimisation, and data subject rights.

However, the GDPR provides limited exemptions (Art. 2(2)) for activities outside the scope of EU law, common foreign and security policy activities. Furthermore, where data is processed for the purposes of law enforcement, the Law Enforcement Directive (LED) 2016/680[7] and its national implementing laws govern data processing instead – a distinction with significant practical implications for HiPSTer deployment scenarios.

EU Member States may establish supplementary legislation on civilian data protection. Lithuania's Law on Legal Protection of Personal Data (LPPD)[8], for instance, supplements the GDPR by clarifying supervisory authority roles, national

[4] https://www.echr.coe.int/documents/d/echr/convention_ENG

[5] https://www.opensocietyfoundations.org/uploads/c8c58ad3-fd6e-4b2d-99fa-d8864355b638/the-concept-of-chilling-effect-20210322.pdf

[6] https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

[7] https://eur-lex.europa.eu/eli/dir/2016/680/oj/eng

[8] https://www.e-tar.lt/portal/lt/legalAct/TAR.5368B592234C/asr

enforcement procedures, and specific processing conditions at the national level.

Law enforcement directive and the civilian/security distinction

Many LED requirements mirror those established in the GDPR, though with specific exemptions applicable to law enforcement contexts – for example, Art. 20 LED also establishes the privacy-by-design principle. The LED applies when systems are used by competent authorities for purposes defined in Art. 1(1): prevention, investigation, detection, or prosecution of criminal offences; execution of criminal penalties; and safeguarding against threats to public security.

Special categories of data – political opinions, religious beliefs, and similar sensitive attributes – frequently arise in OSINT-based analysis. When competent authorities process such data, Art. 10 LED permits such processing only where authorised by EU or national law, necessary to protect vital interests, or where the data subject has manifestly made the data public. The GDPR imposes a more complex and restrictive regime for processing special categories; the LED, by contrast, affords law enforcement authorities greater operational flexibility. As a directive requiring national transposition, member states may establish additional rules. Lithuania's Law on Legal Protection of Personal Data Processed for the Prevention, Investigation, Detection or Prosecution of Criminal Offences, the Enforcement of Penalties or for National Security or Defence Purposes[9], for instance, permits special category processing when based on international obligations in the domains of national security or defence (Art. 8(4)). While there is greater operational leeway for data processed for law enforcement purposes, the applicability of LED and the national Lithuanian law shall still ensure that the right to the protection of personal data and other fundamental rights and freedoms of natural persons are respected.

This regulatory architecture creates a fundamental distinction that HiPSTer-like systems must navigate. A system developed and deployed solely by a civilian technology company falls fully within GDPR and national data protection laws, with automated decision-making and related personal data processing subject to heightened scrutiny. However, systems deployed by law enforcement authorities may access the LED's more permissive framework. National security authorities, while not directly falling under the LED, typically enjoy a comparably flexible framework as related to their functions – as explicitly provided in the Lithuanian law referenced above, which extends these provisions to national security and defence contexts. Recent scholarship highlights the legal uncertainties this distinction creates: Tzanou and Vogiatzoglou (2025) demonstrate through case studies of Palantir data-mining software and Pegasus spyware surveillance that the current "controller-focused approach" – which grounds EU law applicability on the activities of private entities rather than the protection of data subjects – produces fragmented outcomes that emerging surveillance technologies readily exploit. They propose a "data-subject-centric model" that would yield greater legal certainty for systems operating across civilian and security contexts.

For HiPSTer, the practical implication is clear: while privacy-by-design and related data protection requirements must be embedded from system inception regardless of deployment context, operational deployment by law enforcement and national security authorities enables access to regulatory frameworks better suited to hybrid threat detection requirements.

AI Act classification and exemptions

Given HiPSTer's AI components for real-time hybrid threat detection, the EU Artificial Intelligence Act (Regulation 2024/1689)[10] constitutes a critical regulatory layer. The AI Act exempts systems placed on the market, put into service, or used exclusively for military, defence, or national security purposes, regardless of the entity type (Art. 2(3)). Where this exemption does not apply, compliance obligations must be assessed.

The AI Act categorises systems by risk level: unacceptable risk (prohibited), high risk, limited risk, and minimal risk[11]. HiPSTer-like systems deployed in law enforcement or public security contexts will likely qualify as high-risk under Annex III, triggering substantial compliance requirements including, among others, human oversight, fundamental rights impact assessments, data governance obligations, and technical documentation.

However, Erdogan (2024) identifies a significant "national security loophole" at the intersection of the LED and AI Act. Through case studies of Clearview AI, SyRI, and BriefCam, she demonstrates that transparency obligations fragment across regulatory instruments: the AI Act addresses transparency for providers and deployers, while the LED governs data subject rights – yet neither framework adequately addresses scenarios in which AI systems operate across both civilian and security domains. This gap is particularly relevant for HiPSTer-like systems, which may be developed by civilian research institutions but deployed by national security authorities, potentially falling between regulatory frameworks in ways that diminish transparency and accountability.

Intelligent legislation

As HiPSTer-like systems are designed for use by intelligence agencies, national intelligence legislation governs their

[9] https://e-tar.lt/portal/lt/legalAct/TAR.299D835159BE/asr

[10] https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

[11] https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

operational deployment. Lithuania's Law on Intelligence[12] illustrates the typical framework: it establishes the legal basis, tasks, principles, rights, obligations, and oversight mechanisms for intelligence activities. Intelligence authorities may obtain information from open sources, public information systems, and natural and legal persons, subject to statutory safeguards and, where applicable, judicial authorisation.

Art. 16^1 specifically addresses the processing of personal data in intelligence contexts. Intelligence agencies may process personal data, including special categories, for national security or defence purposes. When processing data across different subject categories, agencies must maintain separation to the extent possible. Critically, data subject rights – access, rectification, erasure, and restriction of processing – may be limited in whole or in part where their exercise would reveal operational methods, tactics, or damage intelligence activities, provided such limitations are necessary and proportionate.

For HiPSTer-like systems deployed by intelligence agencies, these provisions enable reliance on statutory exemptions – including limitation of data subject rights – to the extent processing serves intelligence purposes. Similar frameworks exist across EU member states, though specific provisions vary with national transposition choices.

Criminal law frameworks

An accurate HiPSTer-like system design requires an analysis of criminal law frameworks that define hybrid threat-relevant offences. In Lithuania, the Criminal Code[13] includes provisions directly relevant to hybrid threat detection: Art. 172 (interference with voting rights in elections or referendums), Art. 285 (false reporting of public dangers or disasters), and Art. 287 (threats against civil servants or public administration officials).

These legal standards carry important implications for automated content analysis. Systems must be designed for context-sensitive assessment rather than keyword-based detection alone. The distinction between protected speech and criminal conduct often turns on intent, context, and circumstances that require human judgment. This reinforces a design principle evident throughout HiPSTer's architecture: borderline cases require mandatory human review, and system outputs should support rather than replace prosecutorial and judicial decision-making.

Digital services act and platform obligations

At the EU level, the Digital Services Act (DSA)[14] provides relevant background for HiPSTer-like systems, though its primary addressees are online platforms rather than law enforcement. The DSA requires very large online platforms and search engines to identify and mitigate systemic risks linked to disinformation spread, including risks to civic discourse, electoral processes, and public security (Arts. 34–35)[15]. Although the DSA does not directly address hybrid disinformation campaigns or law enforcement functions, its risk-assessment and mitigation logic may inform HiPSTer-like system development – particularly where such systems or their components might be used to support civilian platform compliance rather than security operations.

National information and media law

Member states maintain additional legal frameworks that affect the deployment of HiPSTer-like systems. Lithuania's Law on Public Information[16] establishes the basis for monitoring, assessing, and responding to harmful information activities, including foreign-state-driven propaganda and coordinated disinformation campaigns. Art. 19 prohibits the dissemination of intentionally false information harmful to national security or public order, Arts. 24–26 require transparency in media ownership to reduce risks of foreign manipulation.

Following Russia's aggression against Ukraine in 2022, various amendments enabled targeted, temporary restrictions on rebroadcasting or online distribution of foreign state media disseminating war propaganda or systematically supporting aggression. Within this framework, HiPSTer-like systems can support lawful implementation by identifying content from restricted sources, detecting coordinated disinformation narratives, and providing structured analysis for regulatory and judicial decision-making.

The Lithuanian example illustrates a broader point: while national law establishes legal bases for hybrid threat response, the applicable regulatory framework ultimately depends on which actors develop and deploy HiPSTer-like systems, as well as the operational context of their use.

Cybersecurity requirements

HiPSTer-like systems must ensure appropriate cybersecurity levels corresponding to their operational environment. The

[12] https://www.e-tar.lt/portal/lt/legalAct/TAR.1881C195D0E2/MLwGUKbvnw.

[13] https://www.e-tar.lt/portal/lt/legalAct/TAR.2B866DFF7D43/asr.

[14] https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng

[15] https://www.law.kuleuven.be/citip/blog/can-the-european-union-counter-hybrid-threats-via-legislative-measuresthe-example-of-the-digital-services-act/

[16] https://www.e-tar.lt/portal/lt/legalAct/TAR.065AB8483E1E/asr.

NIS2 Directive[17] does not directly impose cybersecurity requirements on law enforcement or national security agencies, but may affect HiPSTer-like systems deployed within environments involving designated cybersecurity entities. At the national level, Lithuania's Cybersecurity Law[18] and its implementing regulations establish additional requirements that may apply to HiPSTer-like systems; further cybersecurity obligations may arise from legislation governing military and intelligence authorities.

Intellectual property and web scraping considerations

As included further, to collect meaningful intelligence, HIPSTer-like systems engage in web scraping to obtain OSINT/SOCINT. Web scraping, while common in the data-driven economy, is surrounded by extensive debate as to the extent to which such data collection is legal. First, web scraping, which involves personal data, poses the risks to data protection described above. Second, depending on the scraped sources and the HIPSTer-like system's use cases, intellectual property protection requirements should be taken into account. According to the OECD, 'scraping activities can implicate several types of IP and similar rights, including copyright, database rights, trademarks, trade secrets, publicity, and moral rights.' [19] Therefore, in the development of HIPSTer-like systems, intellectual property laws should also be considered. For example, the EU addresses copyright issues related to AI training primarily through two legal instruments – the Directive on Copyright and Related Rights in the Digital Single Market (Directive (EU) 2019/790, DSMD)[20] and the AI Act. While the DSMD permits text and data mining (TDM) of lawfully accessible works subject to rights holders' opt-out pursuant to Article 4(3), the EU AI Act reinforces this framework by requiring general-purpose AI providers to implement copyright-compliance policies, respect TDM opt-outs, and publicly disclose summaries of the data used to train their models.[21]

Soft law and policy frameworks

Beyond legally binding instruments, soft law and policy frameworks provide essential context for HiPSTer-like system development. While these frameworks do not impose operational obligations, they establish the strategic rationale and institutional momentum driving hybrid threat detection capabilities.

National security strategies

Lithuania's National Security Strategy[22] articulates the contemporary threat environment that HiPSTer-like systems are designed to address. The Strategy recognises that threats are no longer clear-cut: the lines between war and peace, foreign and domestic risks, and military and civilian actions are increasingly blurred. Foreign actors employ subtle, covert methods to interfere in political, social, and economic life rather than resorting to overt aggression. Russia's authoritarian and aggressive policies – combining military pressure with hybrid tactics – represent not only a threat to Lithuania but a broader risk to the Euro-Atlantic community (points 13–14). The Strategy explicitly tasks public security institutions with developing capabilities to counter hybrid threats in crisis and emergency situations (point 36(5)).

Lithuania's Strategic Communication Coordination Framework (Government Resolution No. 955)[23] operationalises these priorities by establishing institutional responsibilities for threat prevention and crisis management, coordinating information threat monitoring and assessment, responding to information incidents and sustained information pressure, and planning strategic communication campaigns.

The Military Strategy of the Republic of Lithuania[24] frames national defence as a comprehensive response to both military and non-military threats. Recognising that adversaries employ information operations, cyber-attacks, and other hybrid tactics alongside conventional force, the Strategy emphasises preparedness for complex hybrid environments in which the military, informational, cyber, and civil dimensions are closely interrelated. Civil-military cooperation and interoperability between armed forces and civilian institutions are identified as priorities, directly supporting initiatives such as HiPSTer that aim to improve hybrid threat detection, response coordination, and societal resilience.

EU and NATO frameworks

At the European level, the EU Joint Framework on Countering Hybrid Threats[25] establishes coordinated approaches,

[17] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02022L2555-20221227

[18] https://www.e-tar.lt/portal/lt/legalAct/5468a25089ef11e4a98a9f2247652cf4/bjhGPyIkxw

[19] https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf p. 9

[20] https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng

[21] https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf p. 37-38

[22] https://www.e-tar.lt/portal/it/legalAct/TAR.2627131DA3D2/asr.

[23] https://www.e-tar.lt/portal/lt/legalAct/4d789fb0eb8511eaa12ad7c04a383ca0.

[24] https://kam.lt/wp-content/uploads/2022/03/karine-strategija-LT-2016.pdf.

[25] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52016JC0018.

including mechanisms for recognising hybrid threats; improving awareness through strategic communication; building resilience across critical infrastructure, public health, food security, and cybersecurity; crisis prevention, response, and recovery; and deepening EU-NATO cooperation.

NATO's Strategic Concept[26] identifies hybrid threats as a core and persistent challenge to Alliance security. Point 27 commits NATO to investing in capabilities to prepare for, deter, and defend against the coercive use of political, economic, energy, information, and other hybrid tactics by states and non-state actors. Significantly, the Strategic Concept affirms that hybrid operations against NATO Allies could reach the threshold of armed attack, potentially triggering NATO's collective defence obligations under Article 5. NATO accordingly prioritises partnerships to counter hybrid challenges and maximise cooperative strategies with the EU.

These policy frameworks collectively establish the strategic necessity for HiPSTer-like systems. While they do not prescribe technical specifications or operational requirements, they define the threat landscape, institutional responsibilities, and cooperation mechanisms within which such systems must operate. The convergence of national, EU, and NATO strategic priorities around hybrid threat detection provides both mandate and legitimacy for the capabilities HiPSTer is designed to deliver.

NATO's Artificial Intelligence Strategy[27] sets out the Alliance's approach to responsible AI development and use in defence and security contexts. The Strategy recognises that AI is transforming defence and security globally, creating opportunities to strengthen technological capabilities while simultaneously enabling threats to become faster, more complex, and harder to predict. AI is expected to shape nearly all areas of NATO's work – collective defence, crisis management, and cooperative security. HiPSTer-like systems align directly with this policy framework; the Strategy's principles of responsible use will increasingly inform the design requirements for national security-dedicated AI systems.

Current practices in regulatory compliance

The HiPSTer project aims to develop an advanced software solution that closely interacts with personal data and falls within strict GDPR and AI Act regulations. We continue this narrative by identifying current compliance practices, requirements, and threat definitions across EU frameworks that shape system design.

EU policy implementation and regulatory interplay

EU policy implementation documents increasingly address the intersection and coordination challenges among major digital laws covering cybersecurity, data protection, and AI ethics. The European Parliament Study on the Interplay between the AI Act and the EU Digital Legislative Framework (2025a) stresses that the AI Act requires specific cybersecurity measures to protect against AI-specific attacks, including "data or model poisoning" – methods frequently associated with hybrid interference campaigns. The study highlights a regulatory gap: while the AI Act governs content generation (with transparency and labelling obligations), the Digital Services Act focuses on content moderation once it is hosted. The Study recommends harmonising these regimes to prevent "dispersed enforcement" where content generators and hosts apply divergent risk assessments, a concern directly relevant to HiPSTer-like systems operating across generation, detection, and platform interfaces.

Fundamental rights impact assessment

Deployers of high-risk AI systems – including public bodies likely to operate HiPSTer-like systems – must conduct Fundamental Rights Impact Assessments (FRIA) before deployment, thereby creating compliance thresholds that exceed standard product safety requirements. Mantelero (2024) provides the most comprehensive methodological framework for FRIA implementation under Article 27 of the AI Act. His analysis identifies key building blocks: contextual analysis of AI deployment environments, stakeholder mapping, identification of at-risk rights, and documentation of mitigation strategies. Critically, Mantelero emphasises that FRIA is "necessarily an expert-based assessment." While guidelines and templates can facilitate the process, "the impact on fundamental rights cannot be fully automatised" because relevant parameters are "inevitably context-based and grounded on expert knowledge of fundamental rights." This aligns with HiPSTer's design philosophy: the system supports analyst decision-making rather than replacing human judgment in consequential assessments.

Prohibited practices and regulatory boundaries

The AI Act prohibits techniques posing unacceptable risks: subliminal, manipulative, or deceptive techniques that distort behaviour and impair informed decision-making; "social scoring" systems that evaluate individuals based on social behaviour or traits; and biometric categorisation systems that infer sensitive attributes such as race, political opinions, or sexual orientation. HiPSTer-like systems must be designed to avoid these prohibited categories – a constraint that reinforces the framework's focus on detecting coordinated campaign patterns rather than profiling individual users.

The European Parliament Study identifies tension between GDPR and AI Act regarding bias monitoring. The AI Act permits

[26] https://www.nato.int/content/dam/nato/webready/documents/publications-and-reports/strategic-concepts/2022/290622-strategic-concept.pdf.

[27] https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2021/10/22/summary-of-the-nato-artificial-intelligence-strategy.

processing special categories of personal data (e.g., ethnicity) specifically for "bias monitoring, detection and correction" to prevent societal discrimination – creating a complex interplay with GDPR restrictions on sensitive data processing. For HiPSTer, this tension has practical implications: bias auditing of detection algorithms may require processing data that would otherwise be prohibited, necessitating careful legal navigation.

Institutional guidance

The EDPB Guidance on Fundamentals of Secure AI Systems with Personal Data (Glerean, 2025) reinforces AI Act prohibitions on systems that manipulate human behaviour through "subliminal techniques" or exploit vulnerabilities based on age or disability. The Guidance warns against "loss of human agency and autonomy" and "overreliance" on automated systems – concerns that informed HiPSTer's human-in-the-loop architecture.

ENISA's NIS360 2024 Report (2025) emphasises that public administrations, as prime targets for hacktivism and state-nexus operations, should strengthen cybersecurity capabilities leveraging the EU Cyber Solidarity Act [28] and shared service models. ENISA's Threat Landscape (2025) explicitly identifies that state-aligned actors – particularly Russia and China – employ "inauthentic news articles," "fabricated investigations," and AI-driven deepfakes to interfere in EU elections, discredit public institutions, and manipulate public perception. These threat assessments directly validate HiPSTer's operational focus.

The Report on the State of Cybersecurity in the Union (2024) identifies malicious cyber activity as a clear component of wider hybrid threats aimed at destabilising EU society, democracy, and values, particularly through FIMI and disinformation campaigns. This report highlights the psychological and societal dimensions of the threat landscape, noting that hacktivists utilise "Fear, Uncertainty, and Doubt" to amplify the disruptive impact of their operations.

Incidents targeting civil society and the general public represented a notable share of observed events, often involving social engineering and data breaches. The report emphasises the critical need to strengthen "societal capabilities," recommending harmonised national efforts to improve cybersecurity awareness and cyber hygiene among citizens to ensure the population is resilient against these evolving threats.

The European Commission, with the technical input of the AI Office and the European Data Protection Supervisor, has issued several key guidelines to clarify the legal concepts for AI Act implementation, which are particularly important for the HiPSTer project.

First, we have to mention the Guidance for Risk Management of Artificial Intelligence systems (2025). It focuses on the risk of non-compliance with certain data protection principles stressed in the GDPR for which the mitigation strategies that controllers must implement can be technical in nature – namely fairness, accuracy, data minimisation, security and data subjects' rights. It emphasises that "the risk analysis and the risk evaluation aspects are too dependent on the specific processing context and their assessment is better left to each organization in line with their own risk criteria. This means that EU and national institutions of Member States should perform a thorough analysis for each AI system they plan to use in order to also evaluate the likelihood and impact of the risks, and decide on the mitigating measures to address them, as well as on the residual risks."[29]

Guidance on Generative AI, strengthening data protection in a rapidly changing digital era (2025), also stresses the danger of "hallucinations" (the generation of false information) and technical vulnerabilities such as "prompt injection" and "jailbreaks" that could be exploited by malicious actors. It mandates human oversight to mitigate risks to vulnerable populations and to prevent "automation bias".

Finally, the Guidelines on prohibited artificial intelligence practices (2025a) provide an overview of AI practices deemed unacceptable due to their potential risks to European values and fundamental rights. The guidelines specifically address practices such as harmful manipulation, social scoring, and real-time remote biometric identification, among others. These guidelines provide important insights for implementing projects such as HiPSTer. Namely, the Guidelines stress that AI-enabled social scoring risks producing discriminatory outcomes that exclude individuals and groups from societal participation, while enabling surveillance and social control mechanisms fundamentally incompatible with EU values. These prohibitions are designed to protect democratic principles, equality of access to public and private services, and justice across the Union.

CJEU jurisprudence: data minimisation and automated decision-making

The guidelines are designed to ensure the consistent, effective, and uniform application of the AI Act across the European Union. While they offer valuable insights into the Commission's interpretation of the prohibitions, they are non-binding, with authoritative interpretations reserved for the Court of Justice of the European Union (CJEU). In this light, we have to mention that the CJEU has already adopted several important decisions related to the regulatory interfaces of the AI Act and the GDPR, obligations and restrictions for developers of specific systems.

It should be stressed that recent CJEU rulings have likely shaped the practical application and interpretation of the GDPR, with direct implications for AI systems and cybersecurity liability. Therefore, it is important to mention and discuss them in the context of this project. Recently, the CJEU addressed aspects of the project, including automated decision-making, AI ethics,

[28] https://eur-lex.europa.eu/eli/reg/2025/38/oj/eng?utm_source=copilot.com

[29] https://www.edps.europa.eu/system/files/2025-11/2025-11-11_ai_risks_management_guidance_en.pdf?utm_source=copilot.com, page 8.

pseudonymisation, AI model training, GDPR liability and cyber-fines, data minimisation, and profiling, among others.

In the case of "SCHUFA" (2023) the Court explained that a decision based solely on automated processing significantly affects the data subject. These ruling mandates stricter transparency and human oversight for algorithmic scoring and AI systems used in consequential domains.

The CJEU in the recent case of "EDPS v SRB" (2025) clarified that pseudonymised data is not automatically considered personal data for a recipient if they cannot reasonably re-identify the individuals, taking into account all available means. This offers a more nuanced, context-specific view of data use, which is important for AI model development and data sharing within the EU.

The "Deutsche Wohnen" case (2023) significantly addresses GDPR liability and cyber-fines, clarifying that administrative fines under the GDPR can be imposed only for negligent or willful conduct. It rejects the concept of strict liability for GDPR violations, which is highly relevant for assessing liability and financial risk in the event of a cybersecurity breach.

Recent rulings of the Court of Justice of the European Union provide authoritative interpretations of principles directly relevant to HiPSTer-like systems.

In *Meta Platforms Ireland* (2024), the Court reinforced data minimisation, ruling that GDPR limits a social network's use of personal data collected outside the platform for targeted advertising. The Court clarified that data minimisation (Article 5(1)(c) GDPR) requires data to be adequate, relevant, and limited to what is necessary for processing purposes – reflecting the broader principle of proportionality. Storage limitation (Article 5(1)(e) GDPR) requires data retention only for as long as necessary for the purposes of collection or further processing. Controllers must avoid collecting data that is not strictly necessary for processing. The Court characterised the extensive processing of user data for monitoring large portions of online activity as "a serious interference with fundamental rights" under Articles 7 and 8 of the Charter. For HiPSTer-like systems conducting broad OSINT collection, this ruling reinforces the necessity of purpose-specific data collection and defined retention periods.

In *Dun & Bradstreet Austria GmbH* (2025), the Court addressed transparency in automated decision-making. Data subjects may require controllers to explain the procedures and principles actually applied when using personal data by automated means to obtain specific results. This explanation must be provided through relevant information in "concise, transparent, intelligible, and easily accessible form" – constituting the "meaningful information about the logic involved" referenced in Article 15(1)(h) GDPR. This ruling has direct implications for HiPSTer's explainability requirements: the ontology-driven architecture's capacity to trace reasoning pathways from indicators through campaigns to threat assessments aligns with these transparency obligations.

Synthesis: from regulation to implementation

Current regulatory compliance practices reveal a fundamental shift in the EU security landscape – from regulation toward active implementation of a complex "Digital Rulebook" designed to counter interconnected threat domains. Hybrid threats have evolved beyond simple cyberattacks; Foreign Information Manipulation and Interference (FIMI) is now treated as a systemic risk to democracy, frequently synchronised with cyber operations (EEAS, 2026).

Multiple EU documents highlight that Generative AI acts as a force multiplier for these threats, enabling mass production of hallucinations, deepfakes, and toxic output that obscures the truth and disrupts public life. Projects such as HiPSTer are therefore crucial for detecting, attributing, and countering prioritised threats.

However, in light of the EU regulatory framework and current policy practice – and especially recent CJEU decisions reinforcing data minimisation, proportionality, and transparency obligations – HiPSTer-like systems must be implemented in close coordination with the requirements of the AI Act, the GDPR, and other applicable legislation, ensuring the necessary protection of fundamental rights at all project stages. The framework's ontology-driven architecture, human-in-the-loop design, and explainability mechanisms are not merely technical choices but regulatory necessities.

Systemic challenges in regulatory compliance

A recent systematic study on requirements engineering for regulatory compliance of software systems (Kosenkov et al., 2025) identified numerous challenges: abstractness of regulations, conflicts with existing practices, demand for new knowledge, absence of established principles and practices, complexity, resource intensity, enforcement challenges, dynamics of software systems, multiplicity of overlapping regulations, organisational challenges, and regulatory gaps. Of 280 research works analysed, 255 (91%) mentioned at least one of these challenges – indicating that regulatory compliance difficulties are endemic to software development rather than specific to any particular domain.

These challenges and their interdependencies require a comprehensive investigation to derive foundational principles informing systematic methodologies and robust compliance tools. At present, interpretation and application of regulatory frameworks remain predominantly dependent on specialised legal expertise and domain-specific knowledge (Elahidoost, 2024). For HiPSTer-like systems operating at the intersection of AI, data protection, and national security law, this expertise requirement is compounded by the novelty and rapid evolution of applicable frameworks.

Generative AI and compliance automation

Imperial et al. (2025) argue that generative AI has the potential to strengthen regulatory compliance processes. However, significant technical, legal, and organisational challenges remain. Liability and accountability are unclear when GenAI misinterprets regulations or generates incomplete compliance reports. Regulations change frequently, requiring model retraining to avoid obsolescence. Standards and regulations are not yet aligned with GenAI capabilities, and their application continues to require domain expertise. These limitations suggest that while AI may assist compliance processes, it cannot substitute for human legal judgment – a conclusion consistent with HiPSTer's broader design philosophy regarding human-in-the-loop requirements.

Operational guidance: AI bias in law enforcement

Europol's Innovation Lab (2025) has published practical guidance specifically addressing AI bias in law enforcement contexts – directly relevant to HiPSTer-like systems deployed by security authorities. The guide recognises that AI offers opportunities to enhance efficiency and accuracy in law enforcement investigations, border security, criminal justice procedures, and threat detection, but also identifies AI bias as a critical ethical challenge that can lead to discrimination, erosion of trust, and inaccurate outcomes.

The Europol guidance recommends concrete mitigation strategies aligned with AI Act requirements:

- Conduct independent audits of AI systems before and during operational use
- Maintain constant human oversight with the capacity to intervene in automated decisions
- Critically analyse training data with attention to potential discrimination sources
- Promote diversity and ethics in AI development and implementation teams
- Ensure transparency and that system decisions are understandable to operators and the public
- Establish continuous review protocols to assess long-term AI impact

These recommendations reinforce design choices embedded in HiPSTer's architecture: the ontology-driven approach supports explainability and auditability; the human-in-the-loop design ensures intervention capacity; and the multilingual validation methodology discussed later explicitly addresses potential bias across linguistic contexts.

Regulatory alignment as a design principle

The regulatory landscape surveyed in this section – spanning GDPR, LED, AI Act, national intelligence and criminal law, DSA, cybersecurity frameworks, and soft law instruments – does not merely constrain HiPSTer-like systems but actively shapes their design requirements with fundamental rights protection to be implemented at every step. The convergence of transparency obligations (CJEU jurisprudence), fundamental rights impact assessment (AI Act Article 27), bias mitigation requirements (Europol guidance), and human oversight mandates (EDPB guidance) collectively define an architectural template for lawful hybrid threat detection systems.

HiPSTer's ontology-driven semantic integration, explicit uncertainty representation, and human-in-the-loop reasoning are not incidental features but direct responses to this regulatory environment. The framework demonstrates that analytical depth and operational usability can be achieved within – rather than despite – European fundamental rights and data protection frameworks. As hybrid threats continue to evolve and regulatory instruments mature, this alignment between technical architecture and legal requirements will become increasingly critical for operational deployment.

Methodology for the analysis of dynamic threat indicators

The following subsections define the methodological principles and analytical logic of the proposed framework. The specific implementation choices described later serve as illustrative examples and do not constrain the methodology's general applicability.

At its core, the HiPSTer architecture is ontology-driven. The ontological layer serves as the system's semantic backbone, enabling structured representation of actors, narratives, behaviours, indicators, events, and contextual variables. Unlike traditional data schemas, this ontology is not merely descriptive; it encodes analytical relationships and constraints that support automated reasoning, cross-domain reconciliation, and explainability. This approach directly addresses the fragmentation observed in existing OSINT, SOCMINT, and NLP solutions, where analytical outputs remain isolated and require manual interpretation to establish relevance across domains.

Design rationale: from analytical gaps to architectural requirements

The background analysis reveals a consistent pattern: individual analytical capabilities have achieved considerable sophistication within bounded domains, yet hybrid threats systematically exploit the seams between these domains. Botnet detection algorithms achieve 96% accuracy within Domain Name System (DNS) traffic analysis (Suryotrisongko et al., 2022a, 2022b); bot detection architectures demonstrate strong performance on single-platform datasets (Ellaky et al., 2024); entity extraction systems surpass baselines for cyber threat intelligence (Hamzic et al., 2025). However, no current framework provides systematic correlation across these analytical domains – precisely the capability that hybrid threat detection requires.

This gap is not merely technical but architectural. Current approaches treat threat indicators as classification outputs: a post is labelled as bot-generated or not; a domain is flagged as DGA-produced or not; content is scored for propaganda

techniques or not. These binary or probabilistic labels capture local properties but discard the relational structure that makes hybrid threats coherent as campaigns. When a state actor coordinates disinformation narratives with cyber infrastructure exploitation and physical intimidation, the individual indicators may each fall below detection thresholds while the coordinated pattern constitutes an unambiguous threat signal. Detecting such patterns requires not merely aggregating indicators but reasoning over their relationships.

The HiPSTer methodology addresses this requirement through three core design principles derived directly from the identified gaps:

Principle 1: Semantic Integration Over Tool Aggregation. Rather than developing another specialised detection component, the framework provides a semantic layer that connects existing analytical outputs within a unified knowledge representation. This directly addresses the "domain-specific silos" and "analytical fragmentation" limitations identified across all three literature domains. The ontology does not replace bot detectors, propaganda classifiers, or vulnerability scanners; it provides the representational infrastructure for reasoning about relationships between their outputs.

Principle 2: Explicit Relationship Modelling. Hybrid threats manifest through relationships: actors conduct campaigns; campaigns exhibit indicators; indicators map to capability domains; campaigns target populations. Current analytical tools produce isolated detections without encoding these relationships. The ontological framework makes relationships first-class objects, enabling queries such as "which campaigns exhibit both demonization indicators and coordination patterns within 72 hours of event triggers?" This addresses the "absence of cross-domain reasoning frameworks" gap by providing formal semantics for cross-domain correlation.

Principle 3: Capability-Centric Classification. The SOCMINT literature reveals persistent "conceptual confusion between coordination as methodology versus threat indicator" (Rogers & Righetti, 2025). Not all coordination is malicious; not all bot activity constitutes hybrid threat operations. The framework addresses this through capability domain classification using SKOS taxonomies that distinguish operational capabilities (e.g., Dehumanization and Demonization techniques) from behavioural patterns (e.g., coordination signals). This enables analysts to ask not merely "is there coordination?" but "is coordination being employed to amplify demonization narratives targeting specific populations?" – a question that requires both behavioural and semantic dimensions.

These principles translate into specific architectural requirements: (1) formal knowledge representation supporting automated reasoning; (2) hierarchical capability classification aligned with operational threat analysis requirements; (3) data integration pathways connecting real-world assessment data to semantic structures. The following subsections describe how these requirements are realised through a hybrid knowledge graph architecture combining OWL ontologies, SKOS taxonomies, and RDF data integration.

Hybrid knowledge graph architecture

The HiPSTer system architecture is designed to operationalise the integrated methodological principles described earlier, translating them into a scalable, explainable, and deployment-ready analytical environment. Rather than functioning as a monolithic intelligence platform, the architecture reflects a deliberate separation of concerns: data acquisition, semantic integration, analytical reasoning, and decision-support outputs are treated as interdependent but distinct layers. This design choice emerged directly from the project's empirical and comparative analyses, which consistently showed that tightly coupled architectures struggle to adapt to evolving hybrid threat patterns and regulatory constraints.

The HiPSTer framework employs a three-layer architecture that separates concerns while enabling integration (Figure 1):

Layer 1: OWL Ontology Layer – Formal Threat Actor and Campaign Modelling

The Web Ontology Language (OWL) layer provides formal class definitions, object properties, and reasoning constraints for hybrid threat concepts. Core classes include:

hti:HybridThreat – with subclasses hti:DisinformationCampaign, hti:PsychologicalOperation, hti:NarrativeWarfare

hti:ThreatActor – with subclasses hti:StateActor, hti:ProxyOutlet, hti:CoordinatedNetwork

ind:Indicator – with subclasses organised by analytical domain (ind:IdeologyIndicator, ind:FifthColumnIndicator, ind:RadicalizationIndicator, ind:BehaviouralIndicator)

hti:Target – populations, institutions, or leadership groups targeted by campaigns

Object properties encode the relationships that current siloed approaches discard:

hti:conducts – links actors to campaigns they operate

hti:exhibits – links campaigns to indicators they manifest

hti:targets – links campaigns to target populations

hti:amplifies – links proxy outlets or networks to campaigns they propagate

hti:triggeredBy – links campaigns to event triggers exploited for narrative injection

This formal structure enables SPARQL queries that would be impossible with unstructured indicator lists. For example, the query "identify all campaigns where State Actors conduct operations that exhibit both Dehumanization indicators and Coordination patterns" becomes a tractable graph traversal rather than manual analyst correlation.

Layer 2: SKOS Taxonomy Layer – Hierarchical Capability Classification

The Simple Knowledge Organization System (SKOS) layer provides hierarchical classification of threat capabilities aligned with the HiPSTer project's 12 functional domains derived from empirical use case analysis and expert consultation:

Communication Metadata
Language and Phrasing
Symbolism and Imagery
Narratives and Storytelling
Community and Engagement Patterns
Behavioural Patterns and Activity Trends
Visual and Aesthetic Cues
Cultural References and Symbolic Associations
Social Media Metadata and Geolocation
Content Structure and Framing Strategies
Dehumanization and Demonization
Anomalies and Psychological Patterns

Each domain contains detailed functionality concepts organised through SKOS broader/narrower relationships. For example, the Dehumanization and Demonization domain (hyp:DD_Domain) encompasses:

hyp:Othering – out-group boundary construction techniques
hyp:MoralInversion – guilt transfer and attribution reversal tactics
hyp:ExistentialFraming – survival logic normalising extreme measures
hyp:DehumanizingMetaphors – coded lexicon mapping humans to vermin/disease/dirt

The hyp:hasFunctionality property links specific indicators to their capability domain classifications, enabling queries that distinguish coordination employed for demonization from coordination employed for other purposes – addressing the "conceptual confusion" gap identified in the SOCMINT literature.
Layer 3: Data Integration Layer – CSV-to-RDF Mapping

The data integration layer transforms real-world assessment data into Resource Description Framework (RDF) individuals conforming to the ontological schema. Assessment data from the feature-based solution analysis presented earlier provides functionality presence ratings across commercial and research platforms. This data maps to RDF through Comma-Separated Values (CSV) transformation:

- Platform entities linked to functionality assessments via hyp:hasFunctionalityScore
- Indicator instances linked to campaigns via hti:exhibits
- Temporal metadata enabling event-correlation queries

This layer ensures the ontology is not merely a conceptual model but an operational knowledge graph populated with empirical data supporting analyst queries.

Architectural Rationale

The three-layer separation serves specific purposes:

OWL provides formal semantics enabling automated reasoning (e.g., inferring that if Actor A conducts Campaign B, and Campaign B exhibits Indicator C, then Actor A is associated with Indicator C's capability domain)

SKOS provides flexible hierarchical classification that can evolve as threat landscapes change, without requiring ontology restructuring

Data Integration ensures the framework operates on real-world evidence rather than hypothetical constructs

Together, these layers address the architectural gap identified across the literature: they provide the semantic infrastructure for correlating threat indicators across domains that current analytical approaches treat in isolation.

Ontological framework design

The OWL ontology layer provides the formal semantics enabling automated reasoning over hybrid threat concepts. Unlike database schemas that merely structure storage, the ontology encodes domain knowledge as machine-interpretable axioms supporting inference, consistency checking, and semantic queries. This subsection details the class hierarchy, property definitions, and reasoning patterns that operationalise the design principles articulated earlier. The HiPSTer ontological framework is depicted in Figure 1.

[place Picture1. here]

Core class hierarchy

The ontology organises hybrid threat concepts into five interconnected class hierarchies. Figure 2 provides a detailed view of this structure, showing how the core classes – ThreatActor, HybridThreat, Target, Indicator, and EventTrigger – relate through object properties that encode the relationships essential for cross-domain threat reasoning.

[place Picture2 here]

Hybrid threat classes

The root class hti:HybridThreat represents coordinated operations exploiting multiple domains simultaneously. Subclasses distinguish operational modalities:

hti:DisinformationCampaign – deliberate fabrication or manipulation of information

hti:PsychologicalOperation – systematic efforts to influence target audience emotions, attitudes, or behaviour

hti:NarrativeWarfare – sustained contestation over interpretive frameworks and meaning

hti:MisinformationSpread – propagation of false information without deliberate intent (distinguishing organic from orchestrated phenomena)

hti:SocialMediaManipulation – platform-specific exploitation of algorithmic amplification and network effects

This taxonomy addresses the "conceptual confusion" gap identified in the SOCMINT literature by formally distinguishing threat types rather than conflating all coordination under a single label. An analyst can query specifically for DisinformationCampaign instances exhibiting coordination patterns, excluding organic misinformation that may exhibit similar behavioural signatures but lacks orchestration.

Threat actor classes

Threat actors are modelled with subclasses reflecting operational roles within hybrid threat ecosystems:

hti:StateActor – nation-state entities or state-directed operations

hti:ProxyOutlet – media organisations or platforms providing attribution laundering

hti:CoordinatedNetwork – bot networks, troll farms, or organised amplification clusters

hti:InsiderThreat – actors with legitimate access exploited for threat purposes

The ecosystem perspective is critical. Hybrid operations rarely involve single actors; they coordinate state direction through proxy outlets amplified by coordinated networks. The ontology explicitly models these relationships through the hti:amplifies property, enabling queries that trace propagation pathways from origin through ecosystem delivery.

Indicator classes

Indicators represent observable signals that may suggest hybrid threat activity. The class hierarchy organises indicators by analytical domain:

ind:IdeologyIndicator – markers of ideological framing, othering rhetoric, restoration narratives

ind:FifthColumnIndicator – signals of internal subversion, attribution reversal, delegitimisation

ind:RadicalizationIndicator – progression markers from negative sentiment to mobilisation rhetoric

ind:BehaviouralIndicator – coordination patterns, amplification anomalies, temporal clustering

ind:ConspiracyIndicator – conspiratorial framing, existential threat rhetoric, hidden enemy narratives

ind:DehumanizationIndicator – demonizing language, apocalyptic framing, othering that portrays targets as subhuman or existential threats

ind:NarrativeIndicator – victimization framing, hero/villain construction, restoration and persecution narratives

Indicator instances

Beyond the class hierarchy, the ontology is populated with specific indicator instances derived from empirical analysis, as depicted in Figure 3. The class hierarchy and indicator instances can be interactively examined on our GitHub repository (https://github.com/SecOntologyLab/hipster-ontology).

[place Picture3 here]

This organisation reflects the HiPSTer project's empirical indicator development, which identified 143 detailed functionalities across the 12 capability domains through expert consultation with law enforcement and military practitioners. Each indicator class contains specific instances with formal definitions enabling consistent application.

Target classes

Targets represent entities against which hybrid operations are directed:

hti:PopulationTarget – demographic, ethnic, or national groups

hti:InstitutionalTarget – government bodies, media organisations, international institutions

hti:LeadershipTarget – political figures, military commanders, civic leaders

Explicit target modelling enables query-tracking demonisation and escalation against specific groups over time – a capability essential for the experimental use case developed later.

Event trigger classes

Hybrid operations frequently exploit salient events to inject narrative. The hti:EventTrigger class captures events that campaigns leverage – military operations, political developments, natural disasters, anniversaries, and similar moments of heightened public attention. Temporal metadata associated with event triggers enables ±72h correlation windows for detecting event-driven campaign activity, distinguishing opportunistic narrative injection from organic discourse.

Key object properties

Object properties encode the relationships that transform isolated indicator detections into coherent threat narratives. Table 3 summarises the core properties linking the five class hierarchies, and Figure 4 provides the visual representation.

The class hierarchy, and relationships can be interactively examined on our GitHub repository (https://github.com/SecOntologyLab/hipster-ontology).

Table 3. [was removed from here]

[place Picture4 here]

Reasoning patterns

The ontology supports several reasoning patterns essential for hybrid threat analysis:

Transitive Actor-Indicator Association: If ActorA hti:conducts CampaignB and CampaignB hti:exhibits IndicatorC, then ActorA is associated with IndicatorC's capability domain. This enables actor profiling by aggregating indicators across all campaigns an actor conducts.

Ecosystem Pathway Tracing: Given StateActor hti:conducts Campaign and ProxyOutlet hti:amplifies Campaign and CoordinatedNetwork hti:amplifies Campaign, the ontology supports queries identifying complete propagation pathways from state direction through ecosystem delivery.

Temporal Correlation: The hti:triggeredBy property combined with temporal metadata enables detection of event-driven campaigns – content surges within defined windows of salient events that may indicate opportunistic narrative injection rather than organic discourse.

Capability Domain Aggregation: Through hyp:hasFunctionality links to SKOS concepts, the ontology supports queries aggregating indicators by capability domain. This enables questions such as "which actors employ ≥5 distinct Dehumanization techniques?" – the threshold-based profiling demonstrated in the experimental queries.

SKOS taxonomy integration

The Simple Knowledge Organization System (SKOS) layer provides hierarchical capability classification that complements the formal OWL semantics. Where OWL excels at encoding logical relationships and supporting inference, SKOS provides flexible vocabulary management suited to evolving threat landscapes. This separation enables the capability taxonomy to be extended or refined without restructuring the core ontology.

Design Rationale

SKOS was selected over alternative classification approaches for three reasons:

1. *W3C Standard Interoperability:* SKOS builds on RDF and RDFS, ensuring seamless integration with the OWL ontology layer. SKOS concepts can be directly referenced from OWL individuals through object properties, enabling unified SPARQL queries across both layers.
2. *Hierarchical Flexibility:* SKOS broader/narrower relationships support multi-level taxonomies that can capture both high-level capability domains and fine-grained technique variations. New concepts can be inserted at any level without affecting existing classifications.
3. *Concept Scheme Organisation:* SKOS concept schemes group related concepts, enabling modular taxonomy development. The 12 HiPSTer functional domains are implemented as distinct concept schemes that can be independently maintained and versioned.

The Twelve Functional Domains

The HiPSTer SKOS taxonomy depicted in Figure 5 organises threat capabilities across 12 domains derived from empirical use-case analysis and expert consultation. Each domain represents a distinct analytical perspective on hybrid threat operations:

[place Picture5 here]

Dehumanization and demonization domain: detailed structure

The D&D domain (hyp:DD_Domain) serves as the primary focus for our experimental use case and illustrates the taxonomy's hierarchical structure:

```
hyp:DD_Domain (Dehumanization and Demonization)
├── hyp:Othering
│   ├── hyp:CategoricalExclusion ("they are not like us")
│   └── hyp:OutGroupBoundary (identity boundary construction)
├── hyp:MoralInversion
│   ├── hyp:AttributionReversal (target as true aggressor)
│   ├── hyp:GuiltTransfer (blame projection)
│   └── hyp:VictimPerpetrator (victim-perpetrator inversion)
├── hyp:ExistentialFraming
│   ├── hyp:SurvivalLogic ("no choice" legitimisation)
│   ├── hyp:ApocalypticFraming (civilisational threat rhetoric)
│   └── hyp:EliminationistRhetoric (removal/destruction normalisation)
├── hyp:DehumanizingMetaphors
│   ├── hyp:VerminMetaphors (pest/infestation language)
│   ├── hyp:DiseaseMetaphors (contagion/infection language)
│   └── hyp:DirtMetaphors (pollution/contamination language)
└── hyp:MobilisationCues
    ├── hyp:EmpathyRemoval (contempt/disgust triggers)
    └── hyp:ViolenceNormalisation (cruelty permissibility)
```

Taking into account the schema above, Figure 6 depicts the Knowledge Graph for the Dehumanization and Demonization categories.

[place Picture6 here]

Leaf concepts at the terminal nodes – such as hyp:VerminMetaphors, hyp:AttributionReversal, and hyp:EmpathyRemoval – carry formal SKOS properties enabling consistent operational application. Peripheral nodes represent indicator instances used for automated detection. Each leaf concept includes SKOS properties supporting operational deployment:

skos:definition – formal definition enabling consistent analyst application
skos:example – illustrative instances drawn from documented cases
skos:scopeNote – guidance on boundary conditions and edge cases

For instance, hyp:VerminMetaphors carries the definition "Language characterising target groups as pests, parasites, or infestations requiring elimination," with scope notes distinguishing metaphorical pest rhetoric from literal agricultural or public health discourse.

These SKOS properties ensure that leaf concepts can be reliably applied by human analysts. However, the framework also supports automated detection by explicitly linking observable indicators from the OWL ontology layer to their corresponding SKOS functionality concepts. This cross-layer mapping, implemented through the hyp:hasFunctionality property, enables SPARQL queries to trace from detected signals to analytical categories and vice versa.

Indicator-to-Functionality Mapping

The hyp:hasFunctionality property links specific indicators from the OWL layer to SKOS capability concepts. Figure 7 illustrates representative mappings spanning both Dehumanization and Demonization indicators and coordination co-indicators essential for distinguishing orchestrated FIMI operations from organic discourse.

[place Picture7 here]

This mapping enables the critical analytical distinction identified previously: querying not merely for coordination (which may be benign) but for coordination employed to amplify specific capability domains such as demonization. The query "campaigns exhibiting both hyp:DD_Domain indicators and hyp:Coordination patterns" operationalises the methodological requirement to distinguish organic negative discourse from orchestrated FIMI operations.

To support such queries, the SKOS taxonomy does not operate in isolation – it integrates with the OWL reasoning layer

through property chains and inference rules.

Integration with OWL Reasoning

The SKOS layer integrates with OWL reasoning through property chains and SPARQL query patterns:

1. *Capability Domain Aggregation:* Given an actor, traverse hti:conducts → hti:exhibits → hyp:hasFunctionality → skos:broader* to aggregate all capability domains the actor employs across campaigns (Figure 8).
2. *Threshold-Based Profiling:* Count distinct hyp:hasFunctionality targets within a specified domain to identify actors exceeding technique-count thresholds (e.g., ≥5 distinct D&D techniques).
3. *Cross-Domain Correlation:* Query for campaigns that simultaneously exhibit indicators mapping to multiple capability domains (e.g., D&D indicators AND Behavioural coordination patterns), revealing the multi-domain coordination characteristic of hybrid operations.

[place Picture8 here]

Operationalisation: from schema to query

The preceding sections describe the representational infrastructure; this section previews how that infrastructure supports analyst-grade queries. Full query examples and experimental results are presented in the Case Study (Section 4); here, in Figure 9, we summarise the query patterns the architecture enables.

[place Picture9 here]

These patterns demonstrate how the ontological architecture transforms the analytical requirements identified in the literature review – cross-domain correlation, capability-based classification, ecosystem pathway tracing – into tractable SPARQL queries. The Case Study (Section 4) operationalises these patterns against a concrete experimental scenario, applying the D&D domain taxonomy to analyse documented hybrid threat campaigns and validate the framework's detection capabilities.

Case study: demonization escalation in hybrid threat operations

Use case selection and rationale

A recurring feature of contemporary hybrid and influence operations is the engineered escalation of hostility toward target populations, institutions, or leadership. This escalation is typically not an isolated hate speech phenomenon; it forms part of a broader Foreign Information Manipulation and Interference (FIMI) behaviour pattern intended to sow division, undermine democratic support for policy, and constrain decision-making options (EEAS, 2024). The European External Action Service frames FIMI as intentional, strategic, coordinated manipulation that seeks to confuse and polarise audiences – a definition that positions demonization as an operational capability rather than merely objectionable content (*European Union External Action Service (EEAS)*, 2025)

From an experimental standpoint, demonization and hatred escalation (D&D) is particularly valuable for validating the HiPSTer framework because it is inherently cross-layer in nature. D&D campaigns combine: (a) linguistic indicators such as othering rhetoric, dehumanizing metaphors, and existential threat framing; (b) narrative structures including hero/villain framing and moral polarisation; (c) actor tradecraft including amplification networks and information laundering; and (d) governance effects including pressure on institutions and erosion of public trust. The HiPSTer ontology is specifically designed to represent these cross-layer relations through OWL/SKOS semantics rather than treating them as disconnected classification tasks – making D&D an ideal test of the framework's integrative capabilities.

Selected operational war environment: Russia-Ukraine information space

We anchor this case study in the Russia-Ukraine war information environment for three reasons. First, the EEAS has investigated hundreds of related FIMI cases and repeatedly identifies Ukraine as the most targeted country across incidents in its annual reporting (EEAS Annual Activity Reports 2023, 2024, 2025). Second, this environment contains multiple observable strands of demonization-escalation tradecraft that can be mapped to HiPSTer's indicator framework. Third, the multi-domain nature of the operations – coordinating information, cyber, and physical activities – provides the necessary observables to populate and query the knowledge graph.

Empirically documented tradecraft patterns in this environment include attribution reversal and moral inversion (shifting blame to justify escalation), persistent "Nazi/terrorist state" framing (delegitimising Ukrainian forces), event-driven opportunism (exploiting salient incidents at peak attention windows), and networked ecosystem delivery (coordinated propagation via proxy media and bot amplification). Figure 10 maps these tradecraft patterns to HiPSTer's indicator and functionality framework.

Tradecraft-to-framework mapping for the case study

Four empirically documented tradecraft patterns (Figure 10, left column) are mapped to HiPSTer indicators (centre) and SKOS functionality domains (right). Attribution reversal and "Nazi/terrorist state" framing map to the Dehumanization & Demonization domain (hyp:DD_Domain) through moral inversion and othering indicators. Event-driven opportunism is

captured through temporal correlation queries using ±72h event windows (EEAS methodology). Networked ecosystem delivery maps to coordination indicators (IDEO-14, 5COL-13), enabling Query Pattern 2 (Attribution-Supporting Analysis). This mapping demonstrates how real-world FIMI tradecraft instantiates the abstract ontological framework described earlier.

[place Picture10 here]

The tradecraft-to-framework mapping illustrated in Figure 10 demonstrates that HiPSTer can *represent* documented FIMI patterns through its indicator and functionality architecture. However, representation alone is insufficient. The critical test is whether the framework can *reason* over these patterns, detecting when they co-occur, escalate, and exhibit coordination signatures that distinguish orchestrated operations from organic discourse. This leads to the core experimental question, which we now turn to.

Operational question

The experimental question this case study addresses is:

> Can the HiPSTer ontology represent and reason over the progression from "negative narrative" to "demonisation and hatred escalation" as an integrated hybrid-threat pattern – and can it support reproducible, machine-assisted analyst queries that distinguish (1) organic hateful discourse from (2) coordinated FIMI-style demonisation campaigns with identifiable tradecraft?

This question targets precisely where HiPSTer claims to be novel over existing approaches. Earlier in our discussion of current state-of-the-art tools, we found that commercial platforms such as Babel Street and research systems such as STARLIGHT excel at classification tasks within bounded domains. They can label content as "toxic" or "hate speech" and detect bot-like behaviour. However, they cannot systematically correlate these detections across domains to identify coordinated campaigns, nor can they distinguish organic negative discourse from orchestrated influence operations.

HiPSTer's innovation is to encode demonisation as an operational capability domain, linked to actor models and campaign patterns, through the ontological framework described before. This enables reasoning queries that current tools cannot support:

- "Is D&D co-occurring with information laundering and cross-platform mirroring?"
- "Does D&D intensity spike around event-trigger windows in a manner consistent with coordinated FIMI operations?"
- "Is D&D content being carried by a known proxy ecosystem rather than organic community discourse?"

Converting qualitative analyst judgement ("this looks like demonisation used for mobilisation") into a formal, queryable semantic structure is the core contribution this case study validates.

Ontology instantiation

The use case instantiates HiPSTer's core OWL classes and activates a subset of SKOS capability domains. Figure 11 illustrates this mapping, showing how abstract ontological structures populate with concrete entities from the Russia-Ukraine information environment.

The instantiation proceeds across three layers. First, the hybrid threat classes (hti:HybridThreat, hti:DisinformationCampaign, hti:NarrativeWarfare) capture specific campaign instances such as "Aggressor Reversal Q1-2024" and sustained "Nazi state" framing operations targeting Ukraine. Second, the threat actor hierarchy (hti:ThreatActor, hti:StateActor, hti:TrollFarm) models the ecosystem structure through which these campaigns propagate – from state direction through proxy outlets to coordinated amplification networks. Third, indicator classes (ind:IdeologyIndicator, ind:FifthColumnIndicator) instantiate the specific D&D linguistic markers detected in content: IDEO-18 (Dehumanization), IDEO-14 (Coordinated Activity), 5COL-08 (Attribution Reversal), and IDEO-24 (Moral Inversion). Target and event trigger classes complete the instantiation by specifying Ukrainian Armed Forces, leadership, and civilians as targets, and anchoring analysis to event windows including the Avdiivka offensive, EU integration talks, and NATO summits.

Within HiPSTer's twelve SKOS functionality domains, this use case primarily activates five: Dehumanization and Demonization (hyp:DD_Domain) as the primary domain capturing othering rhetoric and existential framing; Narratives and Storytelling for hero/villain framing and moral inversion; Content Structure and Framing Strategies for binary polarisation techniques; Community and Engagement Patterns for echo chamber reinforcement; and Behavioural Patterns and Activity Trends for coordination signals and temporal clustering.

[place Picture11 here]

The psychological mechanism underlying this mapping is consistent with established research: dehumanisation operates by reducing empathy and increasing disgust/contempt toward target groups, thereby weakening moral inhibition against harming them (Kteily & Bruneau, 2017). This makes D&D indicators plausible precursor signals in escalatory influence environments – and explains why their co-occurrence with coordination patterns warrants elevated concern.

Measurement model: operationalising D&D escalation

Rather than treating demonisation as a single classification label, HiPSTer operationalises D&D escalation as a bundle of co-occurring indicators that can be queried individually or in combination. This structured approach enables graduated assessment rather than binary classification.

Indicator Families

The measurement model organises D&D detection across five indicator families, each capturing a distinct psychological mechanism in the escalation pathway:

1. Othering / Out-group Boundary Construction – Markers of categorical exclusion ("they are not like us") without necessarily explicit slurs, mapped through ind:IDEO-18 to hyp:Demonizing_Language.
2. Moral Inversion and Guilt Transfer – Claims that the target is the true aggressor or source of atrocities, captured through ind:5COL-08 (Conspiratorial Language) and ind:IDEO-17 (Hero and Villain Framing).
3. Existential Threat Framing – "Survival" or "no choice" logic legitimising extreme measures, mapped to hyp: Apocalyptic_or_Existential_Framing.
4. Dehumanising Metaphors – Language mapping humans to vermin, disease, or contamination, detected through ind:IDEO-18 with specific lexical patterns.
5. Mobilisation / Permissibility Cues – Rhetoric normalising cruelty or calls for exclusion, captured through ind:IDEO-21 (Emotive and Polarizing Language) and ind:5COL-05 (Rapid Mobilization).

Coordination Co-Indicators

D&D content alone may represent organic negative discourse. The critical analytical distinction between organic hate and coordinated FIMI operations requires co-occurrence with behavioural indicators that signal orchestration:

1. Coordinated Posting Patterns (ind:IDEO-14, ind:5COL-61) – Coordination score threshold > 0.7
2. Bot-like Activity (ind:5COL-56) – Algorithmicity flags indicating automated amplification
3. Cross-platform Mirroring (ind:5COL-60) – Content similarity > 0.85 across platforms
4. Event-trigger Temporal Clustering – Activity concentration within ±72h of salient events (EEAS methodology)
5. Proxy Ecosystem Origination – Source attribution to hti:TrollFarm or hti:ProxyOutlet actor types

Figure 12 illustrates how these two indicator families converge through the ontological framework. Both D&D indicators and coordination co-indicators connect to campaign instances via the hti:exhibits property, while campaigns link to threat actors (hti:conducts) and event triggers (hti:triggeredBy). This relational structure enables the queries demonstrated before: the ontology does not merely classify content but reasons across domains to distinguish coordinated FIMI operations from organic discourse.

[place Picture12 here]

With the measurement model established, we now turn to the experimental workflow that populates and queries this structure. The workflow proceeds through five stages – from evidence collection through semantic reasoning – demonstrating how raw observables transform into analyst-grade assessments.

Experimental workflow

With the measurement model established, we now turn to the experimental workflow that populates and queries this structure. The workflow proceeds through five stages – from evidence collection through semantic reasoning – demonstrating how raw observables transform into analyst-grade assessments. Figure 13 illustrates this pipeline.

[place Picture13 here]

Stage 1: Evidence Collection and Incident Framing

Analysis begins from defined incident windows anchored to salient events – the approach the EEAS explicitly employs in its FIMI incident methodology. Event triggers for this use case include the Avdiivka offensive (February 2024), EU integration talks (March 2024), and NATO summit announcements (April 2024). These events define temporal windows (±72h) within which narrative injection and amplification patterns are assessed.

Stage 2: Content and Behaviour Extraction

Observables are collected across three channels: content items (posts, articles, videos, memes with text and metadata extracted for linguistic analysis), propagation graphs (cross-posting relationships, mirroring patterns, burst timing, amplification chains), and source typology (classification as official outlet, proxy media, unattributed account, or bot-amplified content).

Stage 3: Linguistic Indicator Detection

Multilingual detection (EN/LT/RU/ZH) applies validated classifiers for D&D indicator families. The HiPSTer framework explicitly positions D&D as a linguistically anchored capability requiring cross-linguistic detection – addressing the gap identified in Section 2 where general-purpose tokenizers recognize only 1.62% of domain-specific threat terminology.

Stage 4: Semantic Ingestion (CSV→RDF)

Extracted observables are transformed into RDF individuals conforming to the HiPSTer ontology: ind:LinguisticIndicator instances typed by SKOS concepts under D&D, hti:BehavioralPattern instances capturing coordination, bursts, and mirroring, hti:ThreatActor links where attribution signals exist, and temporal metadata enabling event-correlation queries. This transformation ensures semantic consistency, automated validation through SHACL constraints, and query-ready representation.

Stage 5: Reasoning and Analyst-Grade Queries

The populated knowledge graph supports SPARQL queries that operationalise the experimental question. Four query patterns demonstrate the framework's capabilities – each targeting a specific analytical requirement that existing classification-only tools cannot address.

Query pattern 1: escalation cluster detection

Operational question: Find campaigns where D&D indicators co-occur with moral inversion narratives and coordination signals within ±72h of event triggers.

```
PREFIX hti: <http://l3ce.eu/ontology/hybrid-threat-intelligence#>
PREFIX ind: <http://l3ce.eu/ontology/hipster-indicators#>
PREFIX hyp: <http://l3ce.eu/ontology/hypster-functionalities#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?campaign ?campaignLabel ?eventTrigger ?coordinationScore
    (COUNT(DISTINCT ?indicator) AS ?indicatorCount)
    (GROUP_CONCAT(DISTINCT ?indicatorLabel; separator="; ") AS ?indicators)
WHERE {
  ?campaign a hti:DisinformationCampaign .
  ?campaign rdfs:label ?campaignLabel .
  ?campaign hti:triggeredBy ?event .
  ?event rdfs:label ?eventTrigger .

  ?campaign hti:exhibits ?indicator .
  ?indicator rdfs:label ?indicatorLabel .
  ?indicator ind:hasFunctionality ?func .
  ?func skos:broader* hyp:Dehumanization_and_Demonization .

  ?campaign hti:hasCoordinationScore ?coordinationScore .
  FILTER(?coordinationScore > 0.7)
}
GROUP BY ?campaign ?campaignLabel ?eventTrigger ?coordinationScore
HAVING (COUNT(DISTINCT ?indicator) >= 3)
ORDER BY DESC(?coordinationScore)
```

Query pattern 2: attribution-supporting pattern analysis

Operational question: Identify D&D content originating in proxy outlets that subsequently enters mainstream platforms via coordinated amplification.

```
PREFIX hti: <http://l3ce.eu/ontology/hybrid-threat-intelligence#>
PREFIX ind: <http://l3ce.eu/ontology/hipster-indicators#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?actor ?actorLabel ?actorType
    (COUNT(DISTINCT ?campaign) AS ?campaignCount)
    (COUNT(DISTINCT ?indicator) AS ?ddTechniques)
WHERE {
  ?actor a ?actorType .
```

```
  FILTER(?actorType IN (hti:StateActor, hti:TrollFarm, hti:BotNetwork))
  ?actor rdfs:label ?actorLabel .

  ?actor hti:conducts|hti:amplifies ?campaign .
  ?campaign hti:exhibits ?indicator .
  ?indicator ind:hasFunctionality/skos:broader* hyp:Dehumanization_and_Demonization .
}
GROUP BY ?actor ?actorLabel ?actorType
ORDER BY DESC(?ddTechniques)
```

Query pattern 3: temporal demonization tracking

Operational question: Track monthly demonization intensity against specific target groups, correlated with battlefield or political events.

```
PREFIX hti: <http://l3ce.eu/ontology/hybrid-threat-intelligence#>
PREFIX ind: <http://l3ce.eu/ontology/hipster-indicators#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?month (COUNT(?indicator) AS ?ddIntensity) ?targetGroup
WHERE {
  ?campaign a hti:DisinformationCampaign .
  ?campaign hti:targets ?target .
  ?target rdfs:label ?targetGroup .

  ?campaign hti:exhibits ?indicator .
  ?indicator ind:hasFunctionality/skos:broader* hyp:Dehumanization_and_Demonization .

  ?campaign hti:detectedOn ?date .
  BIND(SUBSTR(STR(?date), 1, 7) AS ?month)
  FILTER(?targetGroup = "Ukrainian Armed Forces")
}
GROUP BY ?month ?targetGroup
ORDER BY ?month
```

Query pattern 4: actor capability profiling

Operational question: Identify actors who employ≥5 distinct D&D techniques with evidence of ecosystem coordination.

```
PREFIX hti: <http://l3ce.eu/ontology/hybrid-threat-intelligence#>
PREFIX ind: <http://l3ce.eu/ontology/hipster-indicators#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?actor ?actorLabel
    (COUNT(DISTINCT ?indicator) AS ?techniqueCount)
    (GROUP_CONCAT(DISTINCT ?indicatorID; separator=", ") AS ?techniques)
WHERE {
  ?actor a/rdfs:subClassOf* hti:ThreatActor .
  ?actor rdfs:label ?actorLabel .
  ?actor hti:conducts ?campaign .

  ?campaign hti:exhibits ?indicator .
  ?indicator ind:indicatorID ?indicatorID .
  ?indicator ind:hasFunctionality/skos:broader* hyp:Dehumanization_and_Demonization .
}
GROUP BY ?actor ?actorLabel
HAVING (COUNT(DISTINCT ?indicator) >= 5)
ORDER BY DESC(?techniqueCount)
```

Query results summary

Figure 14 summarises the results obtained when executing these four query patterns against the populated knowledge graph. The results demonstrate that HiPSTer can transform qualitative analyst intuitions – "this campaign looks coordinated and is escalating demonization" – into formal, reproducible, machine-assisted assessments grounded in explicit indicator definitions and relationship structures.

[place Picture14 here]

Interpretation and limitations

The experimental results validate HiPSTer's core claim: the ontological framework enables reasoning capabilities unavailable in classification-only approaches. The four query patterns collectively demonstrate that qualitative analyst intuitions – "this campaign looks coordinated and is escalating demonization" – can be transformed into formal, reproducible, machine-assisted assessments grounded in explicit indicator definitions and relationship structures.

Three findings warrant particular emphasis. First, the co-occurrence of D&D indicators with coordination signals and event-

trigger timing (Query 1) produces a detection pattern inconsistent with organic discourse – this is the key discriminator between hateful content and orchestrated FIMI operations. Second, the ecosystem pathway tracing (Query 2) provides attribution-supporting evidence by revealing propagation from state actors through proxy outlets to amplification networks – a capability no classification-only tool can replicate. Third, the threshold-based actor profiling (Query 4) identifies sophisticated operators employing multiple D&D techniques, enabling prioritised analyst attention.

Limitations

Several limitations bound the experimental claims:

1. Indicator validation: The D&D indicator families are operationalised through existing classifiers and lexicons. Detection accuracy varies by language, with stronger performance for English and Russian than for Lithuanian or Chinese. False negatives may miss novel evasion techniques; false positives may flag legitimate criticism.
2. Coordination thresholds: The 0.7 coordination score threshold and ±72h event windows are heuristic values derived from EEAS methodology. Optimal thresholds may vary by context and require empirical calibration.
3. Attribution inference: The queries support attribution assessment but do not provide definitive attribution. State actor involvement is inferred from ecosystem patterns rather than direct evidence.
4. Temporal coverage: The populated knowledge graph reflects a bounded observation period (February–May 2024). Longitudinal patterns require sustained data collection beyond the experimental window.
5. Generalisability: The case study focuses on the Russia-Ukraine information environment – the most documented FIMI scenario. Application to other contexts (e.g., election interference, public health misinformation) would req uire additional calibration and validation of indicators.

Despite these limitations, the case study demonstrates that the HIPSTer ontological framework addresses the capability gaps identified across current OSINT, SOCMINT, and NLP research: systematic cross-domain correlation, capability-based classification, and reproducible analyst-grade queries over hybrid threat patterns. The implications of these findings for operational deployment and future research are discussed below.

The evaluation presented in this paper is intentionally scoped to reflect the nature of the proposed contribution and the operational characteristics of hybrid threats. Validation focuses on three aspects: (i) the internal coherence and consistency of the ontological framework and indicator structure; (ii) the behaviour of selected analytical components, including natural language processing modules, under realistic data conditions; and (iii) the plausibility and interpretability of cross-domain correlations when applied to a historically grounded influence scenario. The evaluation does not aim to measure end-to-end system accuracy, predictive performance, or attribution correctness against a definitive ground truth. Such metrics are often unavailable or misleading in hybrid threat contexts, where intent, coordination, and impact are deliberately obscured. Instead, the evaluation assesses whether the proposed methodology supports earlier, more interpretable, and more structured analytical reasoning compared to siloed approaches, while remaining compatible with legal, ethical, and governance constraints. The following section provides empirical validation of the multilingual detection capabilities that populate the ontological framework with real-world observations.

Empirical validation: multilingual hate speech detection

The ontological framework's operational utility depends critically on automated threat indicator detection capabilities that populate the knowledge graph with real-world observations. This section presents empirical validation of multilingual hate speech and toxic content detection – linguistic indicators that directly populate the SKOS taxonomy's Dehumanization and Demonization domain (Domain 11) – demonstrating the framework's capacity to process threat indicators across the European linguistic landscape and adversary languages.

Detection Architecture and Methodology

The detection system implements a hybrid two-tier classification architecture developed in collaboration with Kaunas University of Technology (KTU):

Tier 1 - Semantic Representation: Sentence transformer models (Potion-base-2M multilingual embeddings) convert raw social media text into 64-dimensional dense semantic vectors that capture meaning-level features, robust to lexical variation, adversarial obfuscation, and cross-lingual contexts. This embedding layer enables language-agnostic threat detection, which is critical for EU-wide deployment, where hybrid threats operate simultaneously across multiple European information spaces.

Tier 2 - Hybrid Classification: Dual detection mechanisms address complementary operational scenarios. Supervised learning via CatBoost gradient boosting optimises the detection of known threat patterns represented in training data. Anomaly detection via HBOS (histogram-based outlier scoring) identifies novel threat narratives and zero-day information operations not represented in historical training data, supporting early warning capabilities for emerging adversarial tactics.

The system processes text through systematic preprocessing (encoding normalisation, URL removal, punctuation standardisation, and emoji conversion) before embedding generation and classification, achieving computational efficiency

suitable for operational deployment on standard cloud GPU infrastructure.[30]

Integration and operational implications

Ontological Integration Workflow

The critical innovation lies not merely in classification performance but in integration with the semantic framework described in Section 6. Detected threat indicators populate the knowledge graph as *LinguisticIndicator* instances linked to broader threat actor profiles and campaign patterns:

1. *Detection:* ML pipeline identifies toxic content in social media stream (e.g., hate speech targeting ethnic minorities)
2. *Semantic annotation:* Detection result instantiated as LinguisticIndicator in knowledge graph with metadata (confidence score, linguistic features, timestamp, source platform)
3. *Relationship inference:* SPARQL reasoning queries link indicator to known ThreatActor profiles exhibiting similar BehavioralPattern signatures
4. *Campaign correlation:* Ontology traverses hti:partOf relationships to assess whether indicator constitutes isolated incident or coordinated campaign element
5. *Attribution assessment:* Semantic reasoning compares observed functionality profile against *hyp:hasFunctionality* attributions for state actors, extremist groups, or inauthentic networks

This integration directly addresses the Dehumanization and Demonization domain (SKOS Domain 11), which encompasses linguistic threat indicators including hate speech, ethnic targeting, group-based hostility, and othering narratives – precisely the content classes validated through the multilingual experiments. The validated detection capabilities thus provide the empirical foundation for the semantic querying demonstrated in Section 7.2: threat actors exhibiting dehumanization techniques can be identified automatically rather than through manual content review.

Implications for European Hybrid Threat Detection

The multilingual validation demonstrates several capabilities essential for operational deployment within European contexts:

Adversary Language Coverage: Validated detection in Russian (primary) and Chinese (secondary) addresses the threat actor languages identified in the adversarial context analysis (Section 3.4), enabling monitoring of source-language information operations before translation and amplification into European target languages.

Pan-European Scalability: Performance consistency across Lithuanian (a low-resource Baltic language) and English (a high-resource language) suggests generalisability to other European languages, supporting the cross-border coordination mandated by EU hybrid threat response frameworks.

Computational Efficiency: Processing of 163,187 Russian texts in approximately one minute demonstrates scalability suitable for real-time monitoring of high-volume social media streams. Projected throughput exceeds requirements for national-scale operational deployment while remaining feasible on standard cloud infrastructure.

Reproducibility: All benchmark datasets are publicly available, and the methodology employs open-source components (sentence transformers, CatBoost, standard preprocessing pipelines), enabling independent validation and extension by other European research institutions – supporting the open science principles emphasised in Horizon Europe funding frameworks.

Conclusions

This paper has addressed a persistent methodological gap in hybrid threat analysis: the absence of a coherent framework capable of integrating heterogeneous indicators across informational, behavioural, psychological, and technical domains without forcing premature attribution or oversimplified classification. Through the **HiPSTer** project, we explored this problem not as a tooling deficit, but as a sense-making challenge rooted in fragmentation, uncertainty, and operational constraints.

The central contribution of this work is methodological. We proposed and detailed an ontology-driven, layered analytical framework that structures hybrid threat detection as a process of progressive interpretation rather than binary decision-making. By combining empirical case reconstruction, state-of-the-art analysis, comparative solution assessment, and formal semantic modelling, the framework provides a systematic way to reconcile weak, dispersed signals into interpretable analytical constructs. This approach explicitly accommodates uncertainty and evolving confidence, which we found to be essential when dealing with adaptive and low-visibility influence operations.

Across the presented system architecture and analytical workflow, the ontology plays a pivotal role. It functions not merely as a data model, but as an integration mechanism that enables cross-domain correlation, traceability, and explainability. Rather than replacing existing OSINT, SOCMINT, or NLP capabilities, the framework orchestrates them within a unified reasoning environment tailored to hybrid threat dynamics. This allows analytical outputs to remain interpretable to practitioners while preserving sufficient depth to capture complex campaign-level patterns.

[30] Infrastructure: Google Colab Tesla T4 GPU. https://colab.research.google.com/notebooks/google.colab.ipynb

The case study on demonisation escalation illustrates how this methodology can be operationalised in practice. The addition of multilingual empirical validation – demonstrating hate speech detection across English, Russian, Lithuanian, and Chinese benchmarks with AUC-ROC scores ranging from 0.672 to 0.931 – provides quantitative evidence that the linguistic indicators populating the Dehumanization and Demonization domain (SKOS Domain 11) can be reliably detected across diverse linguistic contexts and script systems. While the case study results remain observational rather than statistically generalisable, they demonstrate that semantic correlation across narrative, behavioural, and contextual indicators can surface coordinated dynamics earlier than siloed analytical approaches. Importantly, the framework does not claim definitive attribution or predictive certainty; instead, it supports structured analytical judgement aligned with real operational decision-making constraints.

Finally, the regulatory analysis shows that the proposed methodology is compatible with emerging European governance expectations for AI-supported security systems. By embedding explainability, human-in-the-loop reasoning, and proportionality into the analytical design, the framework provides a viable foundation for systems intended to operate within lawful, accountable, and auditable environments.

The methodological contribution presented in this paper is designed to be reusable and partially replicable across different contexts, while recognising the inherent constraints of hybrid threat research. The layered analytical logic, ontology structure, and indicator taxonomy can be reused independently of the specific HiPSTer implementation and adapted to alternative data sources, analytical tools, and operational environments. Full replication of the empirical case analysis is constrained by the sensitivity, availability, and temporal specificity of the underlying data, as well as by ethical and legal restrictions governing access to social media and intelligence-related information. Nevertheless, the framework supports methodological replication through comparable analytical workflows, indicator definitions, and cross-layer integration logic, enabling other researchers and practitioners to reproduce the reasoning process and assess its applicability within their own domains. The OWL ontology, SKOS taxonomy, and SPARQL query patterns are available via the project repository (https://github.com/SecOntologyLab/hipster-ontology), supporting independent evaluation and adaptation.

Limitations

Several limitations should be acknowledged:

- First, while the ontology-driven approach supports rich cross-domain reasoning, its effectiveness depends on the quality, coverage, and maintenance of the underlying knowledge structures. Ontology development remains a resource-intensive task that requires sustained expert involvement.
- Second, while the multilingual validation demonstrates reliable detection across four languages and three script systems, performance varies by language (AUC-ROC 0.931 for Chinese versus 0.672 for Lithuanian), and the case study focuses on a specific class of influence dynamics within the Russia-Ukraine information environment. The adversarial DynaHate benchmark results (63% accuracy, AUC-ROC 0.678) confirm that detection of sophisticated, deliberately challenging hate speech remains an open research problem. Broader validation across additional geopolitical contexts and operational environments is required to assess generalisability.
- Third, although the framework is designed to scale, practical deployment at the national or multinational level introduces challenges related to data availability, interoperability with existing systems, and organisational adoption. These aspects extend beyond methodological design and require institutional alignment and long-term investment.

Multilingual validation results

Rigorous evaluation across four languages and three script systems validates the framework's multilingual capability. Table 4 summarises performance metrics from 10-fold cross-validation on established hate speech and toxic content benchmarks.

Table 4. [was removed from here]

The Chinese Offensive Language Detection (COLD) benchmark demonstrates exceptional performance (AUC-ROC 0.931, substantial agreement κ=0.72), validating robust detection of Mandarin-language threat indicators – directly relevant given the Chinese information operations identified in Section 3.4. Russian toxic content detection (RuToxic, n=163,187) achieves strong discrimination (AUC-ROC 0.777) at scale, addressing the primary adversary language for European hybrid threat scenarios.

Performance on Adversarial Benchmarks

The English DynaHate benchmark merits particular attention as a deliberately adversarial dataset (n=41,144) constructed through human-and-model iterative generation specifically designed to challenge state-of-the-art classifiers. Figures 15 and 16 present ROC and Precision-Recall curves demonstrating detection performance.

[place Picture15 here]

Figure 15 Title + Legend text to be inserted

[place Picture16 here]

Figure 16 Title + Legend text to be inserted

Performance Contextualisation: The 63% accuracy on DynaHate's adversarial examples approaches the 68% reported by the dataset creators using RoBERTa-large models (355M parameters, 1024-dimensional embeddings), while the current approach employs 16× smaller embeddings and a substantially simpler classification architecture. This near-parity validates an efficiency-focused design philosophy: achieving competitive performance with lower computational requirements suitable for operational deployment on commodity hardware.

Operational Flexibility

Precision-recall analysis reveals tunable operating modes enabling diverse stakeholder adoption:

- *High-confidence alerting* (80% precision at 20% recall): Manual review workflows for resource-constrained agencies
- *Balanced monitoring* (68% precision at 60% recall): Semi-automated moderation for platform trust and safety teams
- *Broad coverage* (62% precision at 80% recall): Automated filtering for open-source intelligence researchers

Future work

From a methodological perspective, further refinement of the dynamic threat indicator weighting and confidence-propagation mechanisms is needed. While the current framework supports evolving interpretations and a more formal treatment of uncertainty, integrating probabilistic reasoning more tightly with semantic inference remains an open area.

At the architectural level, future work should focus on extending multimodal integration, particularly for visual, audio, and geospatial data, which are increasingly relevant in hybrid operations. Incorporating these modalities into the ontology without sacrificing interpretability is a non-trivial challenge that warrants dedicated study.

Additional empirical validation is being planned. Applying the framework to a broader set of historical and near-real-time cases, including those outside the European information environment, would strengthen confidence in its robustness and adaptability. Collaboration with operational end-users will be essential in this phase to ensure that analytical outputs remain aligned with practical decision-making needs.

The TRL-4 validation achieved through this research lays the foundation for progression toward operational deployment (TRL 7-9). Development and commercialisation responsibility beyond TRL-4 rests with Novian (2024), a Lithuanian industry partner, with a pathway targeting system prototype demonstration in operational environments within 1–3 years following research completion. This transition will require real-time integration with social media APIs, privacy-preserving deployment architectures, explainable AI dashboards for analysts, and cross-border data-sharing protocols that enable EU-wide coordination without centralising sensitive intelligence.

Data availability

Underlying data

HIPSTer ontological framework

Bružė E, Paskauskas RA, Matulytė R, Sabaliauskaitė G, Lavišius T. HIPSTer — Ontology for Hybrid-Threat Intelligence (OSINT · SoCMINT · NLP). Zenodo. 2025. https://doi.org/10.5281/zenodo.18719337

Licence: CC-BY 4.0

Development repository: https://github.com/SecOntologyLab/hipster-ontology

This archive contains:

- Core OWL ontology (/ontology/hipster-model.ttl)
- SKOS taxonomy for 12 functional domains (/taxonomy/hyp-functionalities.ttl)
- Indicator instance data across 8 indicator categories (/indicators/)
- SPARQL query templates for the four reasoning patterns demonstrated in the Case Study (/queries/)
- Interactive visualisations (/visualisations/)

Files are provided in open, non-proprietary formats (Turtle/TTL, SPARQL, HTML) conforming to W3C Semantic Web standards. All artefacts required to reproduce the ontological reasoning patterns described in the manuscript are

included in the archive.

KTU-Misija-HIPSTer: multilingual hate speech detection pipeline

Bružė E, Paskauskas RA, Matulytė R, Sabaliauskaitė G, Lavišius T. KTU-Misija-HIPSTer: Multilingual Toxic Content Detection Pipeline. Zenodo. 2025. https://doi.org/10.5281/zenodo.18719529

Licence (pipeline code): MIT

Development repository: https://github.com/evavaic/KTU-Misijos-HIPSTer

This archive contains the ML/NLP validation pipeline, preprocessing scripts, and model configurations (licensed under MIT).

Per-language performance results for multilingual toxic content detection across English, Russian, Lithuanian, and Chinese are generated by executing the pipeline on the referenced publicly available datasets, which remain subject to their respective licences.

Figures 15 and 16 (ROC and Precision-Recall curves) are fully reproducible by executing the archived Jupyter notebook research-Misija-HIPSTer.ipynb (MD5: 51dd27bb2b5505af14adf8152a284f86) available in the KTU-Misija-HIPSTer Zenodo archive.

Datasets used for validation

DynaHate (English, n=41,144) File: /data/DynaHate.csv in the KTU-Misija-HIPSTer repository above. Vidgen B, Thrush T, Waseem Z, Kiela D. (2021). Learning from the Worst: Dynamically Generated Datasets to Improve Online Hate Detection. *Proceedings of ACL-IJCNLP 2021*, 1667–1682. https://aclanthology.org/2021.acl-long.132

RuToxic (Russian, n=163,187) File: /data/RuToxic.zip in the KTU-Misija-HIPSTer repository above. Dementieva D, Moskovskiy D, Logacheva V, Dale D, Kozlova O, Semenov N, Panchenko A. (2021). Methods for Detoxification of Texts for the Russian Language. *Multimodal Technologies and Interaction*, 5(9), 54. https://doi.org/10.3390/mti5090054

LtHate (Lithuanian, n=4,995) File: /data/LtHate.csv in the KTU-Misija-HIPSTer repository above. Compiled for the HIPSTer project. It is a binary (yes/no) derivative of *Lithuanian Hate Speech Corpus v.1* (Dataset No. 1: Ethnicity/Nationality/Race), deposited in the CLARIN-LT digital library:

> Butkienė, R.; Dambrauskas, E.; Šukys, A.; Žitkus, V. (2025). *Lithuanian Hate Speech Corpus v.1*. CLARIN-LT digital library. http://hdl.handle.net/20.500.11821/69.

The original corpus is distributed under the CLARIN-LT Public End-User Licence (PUB). LtHate constitutes a format transformation of the original annotated levels into a binary classification and is distributed in accordance with the terms of that licence.

COLD (Chinese, n=37,480) Not hosted in the KTU repository. Licence: Apache 2.0. Deng J, Zhou J, Sun H, Zheng C, Mi F, Meng H, Huang M. (2022). COLD: A Benchmark for Chinese Offensive Language Detection. *Proceedings of EMNLP 2022*, 11580–11599. https://doi.org/10.18653/v1/2022.emnlp-main.796 Canonical repository: https://github.com/thu-coai/COLDataset

Related ontological resources: The 5G Hybrid Threats Ontology, which informs components of the HIPSTer framework, is available at PURL: https://purl.org/5g-hybrid-threats

Extended data: No additional extended data beyond the repository contents described above.

Source data: This methodological paper does not generate new primary empirical data. The case study applies the ontological framework to documented influence operations in the Russia–Ukraine information environment, drawing on publicly available institutional reports from the European External Action Service (EEAS, 2024, 2025, 2026), ENISA (2025), and VIGINUM (2024, 2025). Full citations are provided in the reference list.

Software availability

HIPSTer Ontological Framework

- Source code available from: https://github.com/SecOntologyLab/hipster-ontology
- Archived software available from: https://doi.org/10.5281/zenodo.18719337
- Licence: CC-BY 4.0 (ontological artefacts)

The archived release corresponds to the version used for the analyses described in this manuscript. The repository includes the OWL ontology, SKOS taxonomy, SPARQL query templates, indicator instances, and supporting scripts required to reproduce the semantic reasoning patterns presented in the Case Study.

KTU-Misija-HIPSTer ML/NLP Pipeline

- Source code (ML/NLP pipeline) available from: https://github.com/evavaic/KTU-Misijos-HIPSTer
- Archived pipeline available from: https://doi.org/10.5281/zenodo.18719529
- Licence: MIT (OSI-approved open-source licence)

Key software dependencies

The following software components were used in the development and validation of the framework:

Component	Version	Purpose	Reference
Protégé	5.6.4	Ontology development	https://protege.stanford.edu (RRID: SCR_014278)
Stardog Cloud	latest stable release, (accessed 2025)	Knowledge graph hosting and SPARQL execution	https://cloud.stardog.com
sentence-transformers (Potion-base-2M)	5.1.2	Multilingual semantic embeddings	Reimers N, Gurevych I. (2019). https://doi.org/10.18653/v1/D19-1410
CatBoost	1.2.8	Supervised classification	Prokhorenkova L, et al. (2018). https://doi.org/10.48550/arXiv.1706.09516
HBOS (via PyOD)	2.0.5	Anomaly detection	Goldstein M, Dengel A. (2012). Histogram-based Outlier Score (HBOS): A fast Unsupervised Anomaly Detection Algorithm. *KI-2012: Poster and Demo Track*, pp. 59–63

All dependencies are open-source or publicly accessible tools. Version numbers correspond to those used at the time of validation.

Ethics and consent

Ethical approval and consent were not required.

Competing interests

No competing interests were disclosed.

Grant information

This research was prepared as part of the project “Implementation of Mission-Based Science and Innovation Programs” (Project No. 02-002-P-0001). The project is financed by the European Union through the Recovery and Resilience Facility (2021–2027 programming period, New Generation Lithuania plan), and co-funded by the state budget of the Republic of Lithuania administered via the Research Council of Lithuania (Lietuvos Mokslo Taryba). These combined mechanisms ensure alignment with both EU-level resilience and recovery objectives and Lithuania’s national science and innovation priorities.

The funding received for this project is restricted to supporting research and writing activities; it does not cover publication or printing fees.

Acknowledgements

In preparing this manuscript, we acknowledge the limited use of a generative AI tool (*Microsoft Copilot – no version number provided by the tool*) solely to improve readability, grammar, and language clarity, and to troubleshoot citation manager formatting and bibliography consistency issues (e.g., Zotero). All AI-assisted use occurred under direct human oversight. No generative AI tool was used to generate scientific content, perform conceptual analysis, interpret evidence, determine inclusion or exclusion decisions, or contribute to the theoretical, methodological, or empirical arguments of the manuscript.

All AI-assisted text was reviewed and edited using a human-in-the-loop approach, and the authors take full responsibility for the manuscript’s content, reasoning, and conclusions.

References

Al-Yasiri, J. H., Bin Zolkipli, M. F., Farid, N. F. N. M., Alsamman, M., & Mohammed, Z. A. (2024). A threat intelligence event extraction conceptual model for cyber threat intelligence feeds. In *Proceedings of the 2024 7th International Conference on Internet Applications, Protocols, and Services (NETAPPS)* (pp. 1–8). IEEE.

Arora, A., Arora, A., & McIntyre, J. (2023). Developing chatbots for cyber security: Assessing threats through sentiment analysis on social media. *Sustainability*, *15*(17), 13178. https://doi.org/10.3390/su151713178

Artificial Intelligence Risk Management Framework (AI RMF 1.0). (2023). NIST Trustworthy and Responsible AI, National Institute of Standards and Technology.

Authentic8. (2025). *AI-driven OSINT threat hunting: How AI is transforming cyber defense*. https://www.authentic8.com/blog/ai-driven-osint-threat-hunting-how-ai-transforming-cyber-defense

Avrahami, Z., Zwilling, M., & Hajaj, C. (2025). Leveraging OSINT for Advanced Proactive Cybersecurity: Strategies and Solutions. *IEEE Access*, *13*, 154229–154250. Scopus. https://doi.org/10.1109/ACCESS.2025.3603868

Balogun, A. Y., Ismaila Alao, A., & Olaniyi, O. O. (2025). Disinformation in the digital era: The role of deepfakes, artificial intelligence, and open-source intelligence in shaping public trust and policy responses. *Computer Science & IT Research Journal*, *6*(2).

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), 122–139. https://doi.org/10.1177/0267323118760317

Biagio, M. S., Simoncini, S., La Mattina, E., & Morreale, V. (2024). MARPLE: A framework for social media threat intelligence. In *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)* (pp. 1–6). IEEE. https://doi.org/10.1109/ACDSA59508.2024.10467738

Bimyrzakyzy, A., & Alimzhanova, Zh. M. (2024). Identifying cyberthreats through social media research. *Bulletin of Shakarim University. Technical Sciences*, 42–49. https://doi.org/10.53360/2788-7995-2024-3(15)-7

Bizouarn, E., Abdulnabi, A., & Tan, S. (2023). Leveraging open-source intelligence for automated cyber vulnerability discovery. *Compute*

Blueprint Operational Level Exercise (BlueOLEx). (n.d.). https://digital-strategy.ec.europa.eu/en/news/member-states-and-commission-test-collective-cybersecurity-crisis-response

Borisov, I. (2024). Maskirovka – The art of deception à la Russe. *RMT*, 192–207. https://doi.org/10.55535/RMT.2024.4.12

Cardenas, P., Obara, B., Theodoropoulos, G., & Kureshi, I. (2019). Analysing social media as a hybrid tool to detect and interpret likely radical behavioural traits for national security. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 4579–4588). IEEE. https://doi.org/10.1109/BigData47090.2019.9006259

Charon, P., & Jeangène Vilmer, J.-B. (2021). *Chinese influence operations: A Machiavellian moment*. Institute for Strategic Research (IRSEM), Ministry for the Armed Forces.

Cochran, E. S. (2020). China's "Three Warfares": People's Liberation Army influence operations. *International Bulletin of Political Psychology*, *20*.

Deng, J., Zhou, J., Sun, H., Zheng, C., Mi, F., Meng, H., & Huang, M. (2022). COLD: A benchmark for Chinese offensive language detection. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)* (pp. 11580–11599). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.796

C-203/22 (Dun & Bradstreet Austria GmbH). (2025). GDPR Article 15 (Right of Access) & trade secrets in automated decision-making. https://curia.europa.eu/juris/document/document.jsf?text=&docid=297932&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=8125471

C-252/21 (Meta Platforms Ireland). (2024). GDPR data minimisation and targeted advertising. https://curia.europa.eu/juris/liste.jsf?num=C-252/21

C-307/22 (Deutsche Wohnen). (2023). GDPR liability and administrative fines. https://curia.europa.eu/juris/liste.jsf?num=C-307/22

C-634/21 (SCHUFA). (2023). Automated decision-making & AI ethics (GDPR Art. 22). https://curia.europa.eu/juris/document/document.jsf?text=&docid=282187&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=8119755

Directive (EU) 2016/680 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA. (2016, April 27). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016L0680

Directive (EU) 2022/2555 of the European Parliament and of the Council on measures for a high common level of cybersecurity across the Union (NIS 2 Directive). (2022, December 14). https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng

Dover, R. (2020). SOCMINT: A shifting balance of opportunity. *Intelligence and National Security*, *35*(2), 216–232. https://doi.org/10.1080/02684527.2019.1694132

Dragos, V., Dritsas, E., & Traganitis, A. (2022). A comparative study of machine learning techniques for social media threat detection. *Future Generation Computer Systems*, *128*, 118–131. https://doi.org/10.1016/j.future.2021.09.026

Dragos, V., Forrester, B., & Rein, K. (2020). Is hybrid AI suited for hybrid threats? Insights from social media analysis. In *Proceedings of the 2020 IEEE 23rd International Conference on Information Fusion (FUSION)* (pp. 1–7). IEEE.

Eady, G., Paskhalis, T., Zilinsky, J., Bonneau, R., Nagler, J., & Tucker, J. A. (2023). Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior. *Nature Communications*, *14*, 62. https://doi.org/10.1038/s41467-022-35576-9

Eighth progress report on the implementation of the 2016 Joint Framework on Countering Hybrid Threats. (2025). https://defence-industry-space.ec.europa.eu/system/files/2025-01/SWD_Annual-Progress-Report-2024.PDF

Elahidoost, P. (2024). Towards a tool supported approach for regulatory requirements engineering. In *2024 IEEE 32nd International Requirements Engineering Conference (RE)* (pp. 520–524). IEEE.

Ellaky, Z., Benabbou, F., Matrane, Y., & Qaqa, S. (2024). A hybrid deep learning architecture for social media bots detection based on BiGRU-LSTM and GloVe word embedding. *IEEE Access*, *12*, 100278–100294. https://doi.org/10.1109/ACCESS.2024.3430859

ENISA NIS360 2024 report: A comprehensive look at cybersecurity maturity and criticality of NIS2 sectors. (2024). https://www.enisa.europa.eu/news/enisa-nis360-2024-report

ENISA Threat Landscape 2025. (2025). https://www.enisa.europa.eu/sites/default/files/2025-10/ENISA%20Threat%20Landscape%202025_0.pdf

Erdogan, I. (2024). Diving into the iceberg: Establishing transparency in AI for law enforcement. *European Papers*, *9*(3), 956–977. https://doi.org/10.15166/2499-8249/794

European Commission. (2025a, February). *Guidelines on prohibited artificial intelligence (AI) practices*. https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

European Commission. (2025b, July). *General-purpose AI code of practice*. https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

European Commission. (2025c, November 19). *Digital Omnibus Regulation proposal*. https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-regulation-proposal

European Democracy Shield and EU Strategy for Civil Society pave the way for stronger and more resilient democracies. (2025, November 12). https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2660

European Parliament. (2012). Charter of Fundamental Rights of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A12012P%2FTXT

European Parliament. (2025a). *Interplay between the AI Act and the EU digital legislative framework*. Policy Department for Transformation, Innovation and Health. https://www.europarl.europa.eu/RegData/etudes/STUD/2025/778575/ECTI_STU(2025)778575_EN.pdf

European Parliament. (2025b). Motion for a resolution on hybrid provocations. https://www.europarl.europa.eu/doceo/document/B-10-2025-0437_EN.pdf

European Union Agency for Cybersecurity (ENISA). (2023). *Multilayer framework for good cybersecurity practices for AI*. https://www.enisa.europa.eu/sites/default/files/publications/Multilayer%20Framework%20for%20Good%20Cybersecurity%20Practices%20for%20AI.pdf

European Union Agency for Cybersecurity (ENISA). (2024). *2024 report on the state of cybersecurity in the Union*.

European Union External Action Service (EEAS). (2026). Information integrity and countering foreign information manipulation & interference (FIMI). https://www.eeas.europa.eu/eeas/information-integrity-and-countering-foreign-information-manipulation-interference-fimi_en

Europol. (2025). *AI bias in law enforcement: A practical guide*. Europol Innovation Lab Observatory Report. Publications Office of the European Union. https://www.europol.europa.eu/cms/sites/default/files/documents/AI_bias_in_law_enforcement_-_practical_guide.pdf

Fauziyyah, A. K., Adrian, R., & Alam, S. (2023). Analyzing image malware with OSINTs after steganography using symmetric key algorithm. *SinkrOn*, *8*(2), 818–824. https://doi.org/10.33395/sinkron.v8i2.12266

Federal Ministry of the Interior, Building and Community. (2021). *Cyber security strategy for Germany 2021*. https://www.bmi.bund.de/SharedDocs/downloads/EN/themen/it-digital-policy/cyber-security-strategy-for-germany2021.pdf

Finkelstein, D., Yanovsky, S., Zucker, J., Jagdeep, A., Vasko, C., Jussim, L., & Finkelstein, J. (2025). Information manipulation on TikTok and its relation to American users' beliefs about China. *Frontiers in Social Psychology*, *2*, 1497434. https://doi.org/10.3389/frsps.2024.1497434

Fivecast. (2025, January). *OSINT trends for 2025*. https://www.fivecast.com/blog/osint-trends-for-2025/

Glerean, E. (2025). Fundamentals of secure AI systems with personal data. European Data Protection Board (EDPB). https://www.europarl.europa.eu/RegData/etudes/STUD/2025/778575/ECTI_STU(2025)778575_EN.pdf

Goldstein, J. A., & others. (2024). Disinformation capabilities of large language models. *arXiv preprint arXiv:2311.08838*.

Hamzic, D., Skopik, F., Landauer, M., Wurzenberger, M., & Rauber, A. (2025). Enhancing cyber situational awareness with AI: A novel pipeline approach for threat intelligence analysis and enrichment. *ARES 2025 Workshops*, *15995 LNCS*, 44–62. https://doi.org/10.1007/978-3-032-00633-2_3

Hasanain, M., Ahmad, F., & Alam, F. (2024). Large language models for propaganda span annotation. *Findings of EMNLP 2024*, 14522–14532. https://doi.org/10.18653/v1/2024.findings-emnlp.850

HIPSTer - Ontology for hybrid-threat intelligence (OSINT · SoCMINT · NLP). (2025). https://github.com/SecOntologyLab/hipster-ontology

Huang, T., Yi, J., Yu, P., & Xu, X. (2025). Unmasking digital falsehoods: A comparative analysis of LLM-based misinformation detection strategies. In *Proceedings of the IEEE 2025 8th International Conference on Advanced Algorithms and Control Engineering (ICAACE)* (pp. 2470–2476).

Huang, Y., Chen, Z., & Shu, K. (2023). Large language models for misinformation detection: Opportunities and challenges. *ACM Computing Surveys*, *56*(3), 1–38. https://doi.org/10.1145/3582578

Huang, D., Song, J., & Zhang, X. (2025). Semi-Supervised Social Bot Detection with Relational Graph Attention Transformers and Characteristics of the Social Environment. *Information Fusion*, *118*.

Imperial, J. M., Jones, M. D., & Madabushi, H. T. (2025). Standardizing intelligence: Aligning generative AI for regulatory and operational compliance. *arXiv preprint arXiv:2503.04736*.

Jose, J., & Greenstadt, R. (2024). Are large language models good at detecting propaganda? In *Proceedings of the 5th International Workshop on Cyber Social Threats at ICWSM 2024*.

Kalensky, J., & Osadchuk, R. (2024). *How Ukraine fights Russian disinformation: Beehive vs Mammoth* (Hybrid CoE Research Report 11). The European Centre of Excellence for Countering Hybrid Threats. https://www.hybridcoe.fi/wp-content/uploads/2024/01/20240124-Hybrid-CoE-Research-Report-11-How-UKR-fights-RUS-disinfo-WEB.pdf

Kosenkov, O., Unterkalmsteiner, M., Mendez, D., & Fischbach, J. (2024). Regulatory requirements engineering in large enterprises: An interview study on the European Accessibility Act. In *International Conference on Product-Focused Software Process Improvement* (pp. 204–220). Springer Nature Switzerland.

Kteily, N., & Bruneau, E. (2017). Backlash: The politics and real-world consequences of minority group dehumanization. *Personality and Social Psychology Bulletin*, *43*(1), 87–104. https://doi.org/10.1177/0146167216675334

Li, X., Zhang, Y., & Malthouse, E. C. (2024). Large language model agent for fake news detection. *arXiv preprint*. https://doi.org/10.48550/arXiv.2405.01593

Liu, Y., Xu, S., Li, Y., & others. (2025). Evolution of malicious social bot detection: From individual profiling to group analysis and beyond. *Journal of Social Computing*, *6*(3), 258–284. https://doi.org/10.23919/JSC.2025.0017

LtHate dataset. (2025). GitHub - HiPSTer ontology lab. https://github.com/SecOntologyLab/hipster-ontology

Luceri, L., Giordano, S., & Ferrara, E. (2025). Exposing cross-platform coordinated inauthentic activity in the run-up to the 2024 U.S. election. In *Proceedings of the ACM Web Conference 2025 (WWW '25)*. https://doi.org/10.1145/3696410.3714698

Maltego. (2026). https://www.maltego.com

Mantelero, A. (2024). The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. *Computer Law & Security Review*, *54*, Article 106020. https://doi.org/10.1016/j.clsr.2024.106020

Marchiori, M., Conti, M., & Dragoni, N. (2023). Assessing large language models for vulnerability classification. *IEEE Access*, *11*, 118233–118246. https://doi.org/10.1109/ACCESS.2023.3321147

Meta. (2024). *Adversarial threat report Q3 2024*. Meta Platforms, Inc.

Mothe, J., Roche, M., & Tanguy, L. (2020). Countering hybrid security threats through information analysis. *Information Processing & Management*, *57*(6), 102337. https://doi.org/10.1016/j.ipm.2020.102337

Mykolas Romeris University. (2023). Creation of information security and information threats' detection, analysis, research and education ecosystem (NAAS). https://www.mruni.eu/en/creation-of-information-security-and-information-threats-detection-analysis-research-and-education-ecosystem-naas-nr-01-2-1-lvpa-v-835-03-000-nr-01-2-1-lvpa-v-835-03/

Novian's consolidated revenue increased 2.4% in 2024 to EUR 38.9 million. (2024). https://novian.io/news/novians-consolidated-revenue-increased-2-4-in-2024-to-eur-38-9-million/

Paskauskas, R. A., Bruze, E., Sabalauskaite, G., Matulyte, R., & Lavisius, T. (2026). Hybrid-threat intelligence: A critical review of semantic integration challenges and the HIPSTer ontological framework, *in print*. https://ojs.ukscip.com/index.php/jic

Perrina, G., Caselli, T., & Sprugnoli, R. (2022). AGIR: Automatic generation of intelligence reports from structured threat data. *Natural Language Engineering*, *28*(6), 749–770. https://doi.org/10.1017/S1351324922000176

Project KTU-Misija-HIPSTer. (2025). https://github.com/evavaic/KTU-Misijos-HIPSTer

Qi, P., Yan, Z., Hsu, W., & Lee, M. L. (2024). SNIFFER: Multimodal large language model for explainable out-of-context misinformation detection. *arXiv preprint*. https://doi.org/10.48550/arXiv.2403.03170

Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). (2016, April 27). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504

Regulation (EU) 2022/2065 of the European Parliament and of the Council on a single market for digital services and amending Directive 2000/31/EC (Digital Services Act). (2022, October 19). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065

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

Rogers, R., & Righetti, N. (2025). Coordinated inauthentic behaviour on Facebook? A typology of manufactured attention. *Social*

Media + Society. https://doi.org/10.1177/29768624251369784

Sangher, K. S., Singh, A., & Pandey, H. M. (2024). LSTM and BERT based transformers models for cyber threat intelligence for intent identification of social media platforms exploitation from darknet forums. *International Journal of Information Technology*, *16*(8), 5277–5292. https://doi.org/10.1007/s41870-024-02077-5

Sarhan, H., & Hegelich, S. (2025). There was coordinated inauthentic user behavior in the COVID-19 German X-discourse, but did it really matter? *Frontiers in Communication*, *10*. https://doi.org/10.3389/fcomm.2025.1510144

Shulman, H., & others. (2025). Propaganda by prompt: Tracing hidden linguistic strategies in large language models. *Information Processing & Management*, *62*.

Silver, L., Huang, C., & Clancy, L. (2022). *Across 19 countries, more people see the U.S. than China favorably – but more see China's influence growing*. Pew Research Center.

Social Links. (2025, November). *What is OSINT in 2025?* https://blog.sociallinks.io/what-is-open-source-intelligence-osint-in-2025/

Starbird, K., Arif, A., & Wilson, T. (2020). Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW1), Article 020. https://doi.org/10.1145/3392825

Stardog Essentials. (2025). https://cloud.stardog.com/

STARLIGHT Project - Sustainable autonomy and resilience for LEAs using AI against high priority threats. (2025). https://cordis.europa.eu/project/id/101021797

Suryotrisongko, H., Musashi, Y., & Matsumoto, T. (2022a). Explainable AI for botnet detection using OSINT and DGA analysis. *IEEE Access*, *10*, 45912–45927. https://doi.org/10.1109/ACCESS.2022.3167695

Suryotrisongko, H., Musashi, Y., Tsuneda, A., & Sugitani, K. (2022b). Robust botnet DGA detection: Blending XAI and OSINT for cyber threat intelligence sharing. *IEEE Access*, *10*, 34613–34624. https://doi.org/10.1109/ACCESS.2022.3162588

T-557/20 (EDPS v SRB). (2025). Pseudonymisation and personal data. https://curia.europa.eu/juris/liste.jsf?num=T-557/20

Tzanou, M., & Vogiatzoglou, P. (2025). National security and new forms of surveillance: From the data retention saga to a data subject centred approach. *European Papers*, *10*(3). https://www.europeanpapers.eu/e-journal/national-security-forms-surveillance-data-retention-saga-data-subject-centred-approach

Vacas, I., Medeiros, I., & Neves, N. (2018). Detecting network threats using OSINT knowledge-based IDS. In *2018 14th European Dependable Computing Conference (EDCC)* (pp. 128–135). IEEE. https://doi.org/10.1109/EDCC.2018.00031

VIGINUM. (2024). *Portal Kombat: A structured and coordinated pro-Russian propaganda network*. Secrétariat général de la défense et de la sécurité nationale. https://www.sgdsn.gouv.fr/publications/portal-kombat-un-reseau-structure-et-coordonne-de-propagande-prorusse

VIGINUM - Disinformation, a weapon of war. (2025). https://www.defense.gouv.fr/en/news/disinformation-weapon-war

Wang, G., Liu, P., Huang, J., Bin, H., Wang, X., & Zhu, H. (2024). *Knowcti: Knowledge-based cyber threat intelligence entity and relation extraction*. SSRN. https://doi.org/10.2139/ssrn.4711136

WebIQ Voyager Suite. (2025). https://www.webiq.net

Yadav, A., Kumar, A., & Singh, V. (2023). Open-source intelligence: A comprehensive review of the current state, applications and future perspectives in cyber security. *Artificial Intelligence Review*, *56*(11), 12407–12438. https://doi.org/10.1007/s10462-023-10454-y

***

List of Figures 1 to 16, including titles and legends

Figure 1. Network visualisation of the HiPSTer ontological framework.

The graph depicts the semantic architecture underlying the hybrid threat detection system. Five core class hierarchies are represented as larger nodes: HybridThreat (red, centre), ThreatActor (purple, upper left), Indicator (cyan, lower left), Target (orange, upper right), and EventTrigger (green, lower right). Medium-sized nodes represent subclasses within each hierarchy – for example, DisinformationCampaign, PsychologicalOperation, and NarrativeWarfare as subtypes of HybridThreat; or DehumanizationIndicator and BehavioralIndicator as subtypes of Indicator. Smaller peripheral nodes suggest the 143 detailed indicator instances derived from expert consultation with law enforcement and military practitioners. Edges represent object properties: core relationships (conducts, targets, indicates, exploitedBy) connect the five class hierarchies, while PsychologicalOperation is intentionally positioned on the targets edge, reflecting its defining characteristic as the HybridThreat subtype most directly oriented toward targeting. The triangular linkage between StateActor, ProxyOutlet, and CoordinatedNetwork (purple edges, upper left) models the *amplifies* property that traces ecosystem propagation pathways characteristic of hybrid threat operations. An interactive version is available at https://github.com/SecOntologyLab/hipster-ontology in the README file.

Figure 2. HiPSTer ontology core class hierarchy.

The diagram depicts the five principal class hierarchies and their subclasses: ThreatActor (purple), HybridThreat (red), Target (orange), Indicator (cyan), and EventTrigger (green). Object properties (hti:targets, hti:conducts, hti:amplifies, ind:indicates, hti:exploitedBy) connect the hierarchies, enabling SPARQL queries that traverse from actors through campaigns to observable indicators and exploited events.

Figure 3. Example indicator instances derived from 143 detailed functionalities across 12 capability domains.

The figure illustrates representative instances spanning multiple indicator classes: IDEO-14 (Coordinated Activity) and CONSP-12 (Hidden Enemy Narrative) from the ideology and conspiracy domains; 5COL-08 (Attribution Reversal) capturing fifth-column tactics; RAD-GEN-08 (Dehumanizing Metaphors) and DEHUM-01 (Demonizing Language) from the dehumanization domain; BEHAV-03 (Temporal Clustering) representing behavioural coordination patterns; and NARR-05 (Victimization Framing) from narrative analysis. Each instance includes formal definitions that enable consistent operational application.

Figure 4. Object properties (relationships) linking HiPSTer ontology class hierarchies.

The figure visualises the five core object properties connecting the ontology's class hierarchies: hti:conducts links threat actors to the hybrid threat operations they execute; hti:targets connects operations to affected entities; ind:indicates associates observable indicators with hybrid threat campaigns; hti:exploitedBy captures event triggers within ±72h correlation windows; and hti:amplifies traces ecosystem propagation pathways between threat actors. These relationships enable cross-domain reasoning, which distinguishes HiPSTer from classification-only approaches.

Figure 5. SKOS taxonomy structure for the HiPSTer framework.

The twelve functional domains organise 143 detailed threat indicators derived from empirical use case analysis and expert consultation with law enforcement and military practitioners. Each domain represents a distinct analytical perspective: communication and behavioural metadata (domains 1, 6, 9) capture technical coordination signals; language, symbolism, and cultural references (domains 2, 3, 8) identify ideological markers; narratives, content framing, and dehumanization (domains 4, 10, 11) track influence operation techniques; while community patterns and psychological anomalies (domains 5, 7, 12) reveal coordination structures and radicalisation trajectories. The taxonomy is implemented using SKOS (Simple Knowledge Organisation System), enabling hierarchical classification via skos:broader and skos:narrower relationships, while maintaining interoperability with the OWL ontology layer via skos:exactMatch and skos:relatedMatch mappings.

Figure 6. Network visualisation of the Dehumanization and Demonization domain (hyp:DD_Domain).

This domain serves as the primary focus for the experimental use case, illustrating the SKOS taxonomy's hierarchical structure. The central node branches into five analytical categories: Othering (identity boundary construction), MoralInversion (attribution reversal and victim-perpetrator inversion), ExistentialFraming (survival logic and eliminationist rhetoric), DehumanizingMetaphors (vermin, disease, and contamination language), and MobilisationCues (empathy removal and violence normalisation). Interactive version available at https://github.com/SecOntologyLab/hipster-ontology.

Figure 7. Indicator-to-functionality mapping via the hyp:hasFunctionality property

The diagram illustrates how OWL indicator instances (left column) link to SKOS functionality concepts (right column) through the hyp:hasFunctionality property. D&D indicators – IDEO-18 (Othering Rhetoric), 5COL-08 (Attribution Reversal), CONSP-07 (Existential Threat), RAD-GEN-08 (Dehumanizing Metaphors), and RAD-GEN-12 (Moral Disgust Cues) – map to their corresponding analytical categories within hyp:DD_Domain. Coordination co-indicators, IDEO-14 (Coordinated Activity) and 5COL-13 (Troll Farms), map to hyp:Coordination, enabling the critical analytical distinction: queries can identify not merely the presence of demonization content but its co-occurrence with orchestration signals characteristic of FIMI operations

Figure 8. Property chain traversal for integration of OWL-SKOS reasoning.

The diagram illustrates the traversal path enabling three integration patterns: (1) Capability Domain Aggregation traverses the full chain from actor through campaigns and indicators to aggregate all functionality domains employed; (2) Threshold-Based Profiling counts distinct hyp:hasFunctionality targets to identify actors exceeding technique thresholds; (3) Cross-Domain Correlation queries for campaigns exhibiting indicators mapping to multiple domains simultaneously. The dashed arrow indicates transitive closure via skos:broader*, which recursively ascends the SKOS hierarchy to reach top-level capability domains such as hyp:DD_Domain.

Figure 9. Query pattern summary for Section 3.5 operationalisation.

The four patterns demonstrate how the ontological architecture transforms analytical requirements into tractable SPARQL queries. Pattern 1 (Escalation Cluster Detection) identifies coordinated D&D campaigns near event triggers; Pattern 2 (Attribution-Supporting Analysis) traces ecosystem propagation pathways through proxy outlets and amplification networks; Pattern 3 (Temporal Tracking) correlates demonization intensity with battlefield or political events over time; Pattern 4 (Actor Capability Profiling) identifies actors exceeding technique-count thresholds with coordination evidence. The right column shows key properties traversed in each query. Full SPARQL implementations are presented in the Case Study (Section 4).

Figure 10. Tradecraft-to-framework mapping for the Russia-Ukraine case study.

Four empirically documented tradecraft patterns (left column) are mapped to HiPSTer indicators (centre) and SKOS functionality domains (right). Attribution reversal and "Nazi/terrorist state" framing map to the Dehumanization & Demonization domain (hyp:DD_Domain) through moral inversion and othering indicators. Event-driven opportunism is captured through temporal correlation queries using ±72h event windows (EEAS methodology). Networked ecosystem delivery maps to coordination indicators (IDEO-14, 5COL-13), enabling Query Pattern 2 (Attribution-Supporting Analysis). This mapping demonstrates how real-world FIMI tradecraft instantiates the abstract ontological framework described in Section 3.

Figure 11. Ontology instantiation for the Russia-Ukraine case study.

The diagram maps abstract OWL classes (left column) to concrete use-case entities (centre column) and indicates which SKOS capability domains are activated (right column). Primary mappings to the Dehumanization and Demonization domain are highlighted in orange; secondary activations appear in grey. The five OWL class hierarchies – HybridThreat (red), ThreatActor (purple), Indicator (teal), Target (orange), and EventTrigger (green) – maintain the colour coding established in Section 3. The property chain *hti:conducts → hti:exhibits → hyp:hasFunctionality → skos:broader* (Figure 8) enables traversal from actors through campaigns and indicators to capability domains, supporting the query patterns demonstrated previously.

Figure 12. Measurement model for distinguishing coordinated FIMI operations from organic discourse.

The diagram illustrates how D&D indicator families (left, blue) and coordination co-indicators (right, green) converge through the campaign node (centre, red) via hti:exhibits relationships. Campaigns link downward to threat actors (purple) via hti:conducts and to event triggers (orange) via hti:triggeredBy. The decision logic (bottom) shows the ontological reasoning: when D&D indicators co-occur with coordination signals and event-trigger timing, the pattern indicates a coordinated FIMI operation warranting elevated concern; D&D indicators alone suggest organic hateful discourse requiring standard monitoring. This relational structure – unavailable to classification-only approaches – operationalises the graduated assessment capability central to HiPSTer's design.

Figure 13. Experimental workflow pipeline from evidence collection through semantic reasoning.

The five stages transform raw observables (posts, articles, metadata) into analyst-grade assessments distinguishing coordinated FIMI operations from organic discourse. Stage 1 defines event-triggered temporal windows using EEAS methodology (±72h). Stage 2 extracts content items, propagation graphs, and source typologies. Stage 3 applies multilingual classifiers (EN/LT/RU/ZH) to detect D&D indicator families and coordination signals. Stage 4 transforms extracted observables into RDF individuals with SHACL validation. Stage 5 executes SPARQL queries supporting the four reasoning patterns. The bottom panel shows capabilities enabled by this pipeline that are unavailable in classification-only approaches: cross-domain correlation, ecosystem pathway tracing, threshold-based profiling, and temporal event correlation.

Figure 14. Summary of SPARQL query results demonstrating HiPSTer's ontological reasoning capabilities.

Query 1 (Escalation Clusters) identifies three disinformation campaigns with coordination scores exceeding 0.7, all linked to battlefield or political event triggers within ±72h windows. Query 2 (Attribution Analysis) traces ecosystem pathways from the primary state actor (RF-Alpha) through three proxy outlets to coordinated amplification networks. Query 3 (Temporal Tracking) reveals strong correlation (r=0.84) between demonization intensity spikes and event triggers across the February–May 2024 observation period. Query 4 (Actor Profiling) identifies one actor exceeding the 5-technique threshold, employing eight distinct D&D techniques across the full capability spectrum. Together, these results demonstrate capabilities unavailable in classification-only approaches: systematic cross-domain correlation, capability-based classification, and reproducible analyst-grade queries over hybrid threat patterns. Interactive demonstrations are available at https://github.com/SecOntologyLab/hipster-ontology; the demonstrations use simplified query structures for visualisation purposes, while the formal SPARQL implementations above reflect the complete ontological traversal.

Figure 15. Receiver Operating Characteristic (ROC) curves for DynaHate benchmark.

The 2c potion (CatBoost) achieves AUC-ROC 0.678, substantially exceeding the 1c baseline (AUC 0.528) and the random classifier diagonal.

Figure 16. Precision-Recall Curves (PRC) for DynaHate benchmark.

AUC-PRC of 0.699 demonstrates robust positive-class identification despite the inherent class imbalance challenges in hate speech detection.

TABLES

Table 1. Summary assessment of hybrid threat detection solutions based on expert evaluation across 143 detailed functionalities.

Solution	Strength Areas	Limitation Areas	Overall Coverage
EEAS FIMI	Narrative detection, campaign identification	Technical integration, automation	Highest
Logically	Content analysis, misinformation detection	Cross-platform correlation	High
Storyzy	Source tracking, narrative analysis	Behavioural pattern detection	High
Babel Street	Multilingual monitoring, data aggregation	Semantic reasoning, attribution	Moderate
WebIQ	Darknet coverage, visual analytics	Cross-domain integration	Moderate
VIGINUM	Attribution analysis, network mapping	Real-time automation	Moderate
NAAS	Narrative clustering, sentiment analysis	Cross-platform capability	Moderate
STARLIGHT	AI-driven detection, bot identification	Operational integration (TRL 4-6)	Moderate
EU-NATO Fusion Cell	Coordination mechanisms, information sharing	Technical platform integration	Lower

Table 2. Capability domain coverage across assessed solutions, ranging from well-developed baseline capabilities to minimal presence in emerging analytical areas.

Capability Domain	Coverage Assessment
Communication Metadata	Well developed across most platforms
Language and Phrasing	Well developed; baseline capability
Narratives and Storytelling	Present in most; variable depth
Community and Engagement Patterns	Present; influencer networks well covered
Behavioural Patterns and Activity Trends	Partial; bot detection present, coordination detection limited
Content Structure and Framing Strategies	Partial; misinformation detection present, framing analysis limited
Symbolism and Imagery	Limited; meme detection largely absent
Visual and Aesthetic Cues	Limited across solutions
Cultural References and Symbolic Associations	Limited; requires cultural expertise
Social Media Metadata and Geolocation	Variable; platform-dependent
Dehumanization and Demonization	Minimal; critical gap identified
Anomalies and Psychological Patterns	Minimal; emerging research area

Table 3. Key object properties linking HiPSTer ontology class hierarchies.

Property	Domain	Range	Semantic Function
hti:conducts	ThreatActor	HybridThreat	Links actors to campaigns they operate
hti:amplifies	ThreatActor	HybridThreat	Links proxy/network actors to campaigns they propagate
hti:exhibits	HybridThreat	Indicator	Links campaigns to observable indicators
hti:targets	HybridThreat	Target	Links campaigns to target populations
hti:triggeredBy	HybridThreat	EventTrigger	Links campaigns to exploited events
hyp:hasFunctionality	Indicator	skos:Concept	Links indicators to SKOS capability domains
hti:hasCoordinationScore	HybridThreat	xsd:decimal	Quantifies coordination evidence (0.0–1.0)

Table 4. Multilingual hate speech detection performance (CatBoost classifier, 10-fold CV)

Language	Dataset	Script	n	AUC-ROC	Accuracy	Cohen's	Wall Time
Chinese	COLD	Hanzi	37,480	0.931	86%	0.72	~10 min
Russian	RuToxic	Cyrillic	163,187	0.777	70%	0.30	~1 min
English	DynaHate	Latin	41,144	0.678	63%	0.25	~2.5 min
Lithuanian	LtHate	Latin	4,995	0.672	63%	0.20	~27 sec