Deep Learning Techniques for Dark Web Crawling: A Comprehensive Survey for Threat Intelligence

Asjad Amin
Department of Information and Communication Engineering
The Islamia University of Bahawalpur
asjad.amin@iub.edu.pk

Momna Rehman
Department of Cyber Security and Digital Forensics
The Islamia University of Bahawalpur
momnarehmanibex@gmail.com

Sana Younis
Department of Cyber Security and Digital Forensics
The Islamia University of Bahawalpur
sanayounis175@gmail.com

Kishwar Ishfaq
Department of Cyber Security and Digital Forensics
The Islamia University of Bahawalpur
ishfaqkishwar259@gmail.com

***Abstract*—The dark web is a hub of illegal activities that includes malware distribution, cyber fraud and extremist operations. It has become a vital source of information for cyber threat intelligence (CTI). The survey covers measurement studies, hidden services, marketplaces, forums, and law-enforcement case investigations , providing a thorough evaluation of dark-web research similar to CTI. It focus on recent studies from 2023 to 2025 and investigate the development of dark-web datasets related to cryptomarket, forum, traffic, authorship, and named-entity recognition. The survey also examines the use of deep learning methods for darkweb intelligence extraction like multimodal fusion employing attention-based and transformer architectures, detecting criminal intent, classifying encrypted traffic and detecting anomalies. The discussion covers ongoing challenges such as sparse datasets, oppositional behaviour, data distribution shift, scalability and ethical limits. It highlights the research gaps and future objectives and emphasize the resilient learning under changing threat situations, real-time multimodal intelligence fusion, resilient data collecting, and the incorporation of ethical and legal protections. It can be used by researchers and practitioners to create a deep learning-driven CTI platforms based on dark web intelligence.**

I. Introduction

The dark web is a whole mixture of unseen networks and services,not only the hidden area of the internet, mostly accessible through privacy tools like Tor or I2P. As time passes, it's grown into a complex, heterogeneous ecosystem.In some fields people use it to sefeguard their privacy,while others use it for some illegal activities like cybercrime tools, stolen data, weapons trade, extremist content, and fraud. [1] [2] .Dark-web buying and selling platforms has make it easy for people to Exchange and Manage these activities.They use Operational security measures, escrow based systems, user trust scores, and cryptocurrency Payouts to stay under the Monitoring systems.When a specific platform is disrupted by scam or law enforcement, the users look after the alternative sites that are used and allow the activities to continue. [3] [4].Collecting and studying dark-web data is a huge problem. Websites appear and disappear, the content that is available has a huge variety of languages, some webscraping tools are blocked by different platforms, and monitoring the crypto transactions is challenging for the users [4].As a result, Human-driven threat intelligence techniques are useless. Now the researchers can lean on System-driven, multi-modal CTI pipelines, one of them isadvanced robust crawlers, network analysis, extraction of the advanced content, and cryptocurrency tracing. Deep learning is playing a big role in dark web. It identifies the unknown traffic on tor browser, categorizes out listings and posts, and it can even predict migration of marketplaces and vendors. It will early detect the threat and shortens the response time. [5]. This survey paper highlights a deep learning–powered CTI platform for the dark web. It collects different type of features like data of the network , text, graphs, and cryptocurrency payouts and inputs them into upgraded neural models like transformers and graph networks. The end result shows that the whole system has legal and ethical standards and the analysts get quick, intelligence supported by evidence. [6].

Over the past ten years, and especially between 2022 and 2025,variety of datasets have been obsereved. They allow us to examine the authorship, analyze writting patterns, combine information about darknet markets, arrange network traffic, and detect cybercrimes as they happen. Basically, these datasets are the core component for machine learning, NLP, and threat intelligence systems.

Datasets plays an important role: Darkoob is a dataset which focuses on early classification work, the Agora marketplace for Preemptive CTI measures, and another dataset VeriDark is a structured dataset for authorship verification. This review pull these together to show where dark-web dataset research stands now. [7]. The tor is an anonymous communication network that enabled the expansion of the dark web and its rapidly growing, providing ecosystem that is unseen where illegal activities like malware distribution, financial fraud, weapon sales, and data leakage are conducted by the cybercriminals. Traditional monitoring techniques are struggling to identify and analyzing these types of threats due to encrypted traffic, Frequently changing hidden services etc. To overcome these types of Issues, recent research has started working on the deep

DOI: https://doi.org/10.67535/tsp.000002.018
Published under the ISAISS 2026 Proceedings

Fig. 1. Single-column Dark Web CTI pipeline

learning techniques that are capable to extract complex patterns from darknet network traffic, forums, and marketplaces. [7]. To detecting Unusual traffic, to classify darknet flows, and identifying malicious behaviour some models are used such as CNNs, LSTMs, Transformers, and hybrid architectures.

This survey paper proposes *A Deep Learning–Driven Dark Web Platform for Cyber Threat Intelligence*, which employs darknet data with deep learning–based traffic classification, data analysis, and cyber threat prediction. Recent advancements are build in darknet detection frameworks—such as DarkMor [8], DarkDetect [9], ODTC [10], and DarknetSec [9]—this research will help the researchers to maintain the accuracy, system scalability, and machine based processing of dark-web threat analysis.

II. Bckground And Literature Review

The dark web is a unseen part of the internet where people often share leaked data, hacking tools, and plans for cyberattacks.After that the researchers collect dark web data and turn it into practical datasets. Deep learning models are then trained to read and learn so that they begin to recognize Unusual patterns and upcoming cyber threats. This will help security teams to be known about the further risks. Combining dark web data, datasets, and deep learning, cyber threat detection becomes advanced and more proactive.

A. Dark-Web

In-depth studies of the dark web have been conducted in multiple areas relevant to cyber threat intelligence (CTI). Research focusing on dark-web topology, anonymous services, and data-driven studies has analyzed Tor and hidden services to uncover network structures, usage patterns, and operational characteristics. crucial for CTI, including Major hubs,Network graphs ,Protocol usage and exit-relay risks.ore measurement research highlights that direct participation in Tor enables the collection of usable telemetry, while Successive studies have developed automated frameworks for discovery and tracking .onion services and monitor their lifespans [6], [7], [11]. Dark-web marketplaces (DWMs) have been a central focus due to their role Serving as a platform for illegal commerce. Large-scale analyses quantify trading volumes,Marketplace migration after closures, and supply responses to Practical events such as the COVID-19 pandemic.Research on collective behavior emphasizes the rapid recovery and robustness of marketplaces after takedowns, with AlphaBay and Hansa acting as primary case studies demonstrating coordinated law-enforcement actions and the ecosystem's adjustments [6], [11]. Content-based research on crime typologies in the dark web suggests that DWMs and forums host a variety of illicit services, including opioid and fentanyl distribution, weapon sales, stolen data, carding services, malware and ransomware offerings, as well as extremist propaganda and recruitment. Forum discourse studies underscore mechanisms of trust such as vouching, escrow services, and operational security practices that support these unlawful activities.Data gathering and archival initiatives form the foundation of long-term dark-web research, facilitating the examination of longitudinal phenomena such as market lifecycles and extremist forums. Studies on crawler architecture, archive schema design, and visualization tools provide crucial frameworks for constructing resilient CTI pipelines capable of managing multilingual content, anti-crawling mechanisms, and evidence-preservation requirements. [7], [12], [13].

Machine learning (ML) and deep learning (DL) methods are increasingly used for detecting and predicting activity on the dark web. These methods include encrypted traffic classification, identifying forum and listing intentions using BERT and transformer models, anomaly detection with autoencoders, data generation using GANs, and combining traffic, text, and graph features. Studies point out challenges such as limited datasets, changes in data patterns, and robustness against attacks as major obstacles for real-world use. [5], [14].

Forensic approaches, lessons from takedowns, and law-enforcement actions have been studied through detailed case studies like Silk Road, AlphaBay/Hansa, and White House Market. These studies provide insights into investigative methods, including undercover operations, coordinated seizures, and chain-of-custody procedures, highlighting the importance of cross-border cooperation and careful handling of evidence from anonymized networks. [8], [11], [15].

Finally, ethical and privacy issues in dark-web research show the dual nature of the ecosystem. While the dark web can protect activists and privacy-focused users, it can also enable criminal activity. Researchers suggest creating careful policy frameworks and including legal and ethical controls in CTI platforms to balance these concerns.

The Research identifies alot of research gaps encouraging the development of a AI-driven CTI framework. These include the need for Stable data gathering and linking the entities across marketplaces, blockchain networks, and forums [7];To improve the detection multi input modeling will help to combine the textual, temporal, and graph features. [13]; provides strength against malicious entities. [13], [13]; and representation legal and ethical protective measures, such as data integrity, confidential processing,into active CTI systems. [5].

B. Data-sets

Dark-web datasets research starts with the classification and identification of tasks that focus on identifying illegal content. For example the Darkoob dataset, is used for the deep-learning-based classification of weapons, drugs,bank cards, and data related to illegal currency. [16]. Khan et al. increased this direction by suggesting a proactive threat-intelligence system taught on the Agora marketplace dataset, illustrating the strong performance of models in corrupted darknet forums [17]. Authorship verification has became another important and crucial research area with the introduction of VeriDark.

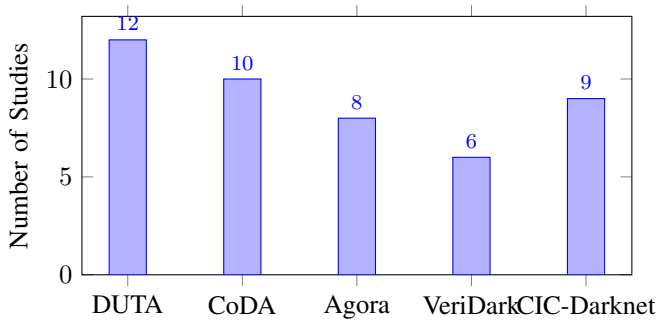

Fig. 2. Usage Frequency of Common Dark Web Datasets

[18]. Linguistic analysis was strengthened through CoDA, a dataset of 10,000 dark-web documents designed for language modeling, topic classification, and behavioral analysis [19]. Similarly, the Evolution cryptomarket dataset provided over half a million forum posts and tens of thousands of listings, enabling analysts to study vendor reputations, illegal product availability, transaction patterns, and marketplace evolution [20].

From 2023 to 2025, a substantial shift occurred toward deep-learning-based intelligence extraction. A 2025 study introduced a comprehensive darknet NER dataset and demonstrated high-accuracy entity extraction using advanced architectures such as GLiNER and UniversalNER [21]. Complementing this, a semi-supervised ensemble model based on ModernBERT was developed to classify illicit activities across the dark web and social platforms using datasets like DUTA and CoDA [22]. DUTA-10K itself remained a widely used onion-service dataset supporting multi-class darknet-site classification [23], while new traffic datasets such as CIC-Darknet2020 enabled Tor/VPN traffic detection and intrusion analysis [24]. CAIDA's darknet scanner data further supported research on global scanning and malicious probing behavior [25], and the SafeSurf 2025 dataset expanded darknet traffic research to include I2P and Freenet [26]. Marketplace-related datasets continued to grow. Multi-market analyses, such as those covering a decade of drug-market activity across 13 major cryptomarkets [27], provided insight into global drug-trade trends. Economic analyses of darknet markets used datasets spanning eight marketplaces to examine pricing, availability, and vendor behavior [28]. The Hydra marketplace was also studied in depth, with researchers releasing structured datasets covering vendors, product categories, and transaction structures [29]. Additional scraping studies, including those related to White House Market [30], contributed to publicly available structured datasets.

Threat-intelligence datasets also expanded significantly. OSINT-based NER corpora such as ND-NER [31], surveys of CTI-NER annotation standards [32], and PAN authorship datasets [33] strengthened the ecosystem of intelligence-oriented corpora. Research on author diarization further advanced authorship-forensic capabilities [34]. NDARC published methodological notes on the construction of cryptomarket datasets, helping standardize future collection efforts [35]. More general studies surveyed the state of dark-web dataset research and identified emerging trends and dataset repositories [36]. Community and industry datasets—including DW-Data GitHub archives [29], threat-intel CSV repositories [30], Zenodo-hosted darknet datasets [33], Cebulka's regional darknet dataset [31], and Bitsight's 2-billion-item underground-activity summary [32]—further expanded the diversity and availability of dark-web intelligence resources.

Together, these 30 works represent the most comprehensive and up-to-date landscape of dark-web datasets, covering crypto markets, forums, threat intelligence, traffic analysis, authorship forensics, linguistic modeling, NER annotation, and market-trend analysis.

C. Deep-Learning

Deep learning has emerged as a dominant approach in dark-web research due to its ability to identify concealed malicious behavior within encrypted or obfuscated traffic. Early works utilizing CNN-LSTM architectures demonstrated significant improvements in darknet traffic categorization, enabling fine-grained detection of command-and-control patterns and malicious web interactions [37], [38]. These sequence-based methods revealed that temporal and spatial traffic features are essential for identifying dark-web activities that mimic legitimate communication environments.

Recent research has used advanced multimodal and attention-based architectures. Yang et al. presented DarkMor, combining local and spatial features with hybrid deep neural networks, greatly improving detection accuracy on darknet datasets [21]. Similarly, Lan et al. proposed DarknetSec, a self-attentive model that allows reliable categorization of encrypted darknet traffic and application identification [4]. Attention-based methods remain widely used in modern research, as they handle mixed darknet datasets more efficiently than traditional feature-based pipelines.

Several studies have focused on real-time darknet threat detection. ODTC, a multimodal Transformer-based model, enabled online classification of darknet traffic with efficient processing and low delay, addressing the scalability limits of earlier offline models [10]. Other frameworks, like ContextualGraph-LLM, combined graph embeddings with large language models to improve understanding of dark-web interactions [39]. These approaches emphasize the importance of combining intelligence from multiple sources, given the growing complexity of cybercriminal networks.

Deep learning has also been applied to analyzing dark-web content. BERT-based models have been used to verify cybercrime intent within dark-web forums [40], GANs for generating synthetic darknet traffic to increase dataset variety [41],and autoencoders for detecting anomalies in Tor traffic [42]. These studies show the Efficiency of neural representations in detecting both traffic-level and content-level malicious activity.

Surveys point out ongoing challenges such as limited datasets, encryption-related feature loss, and weak generaliza-

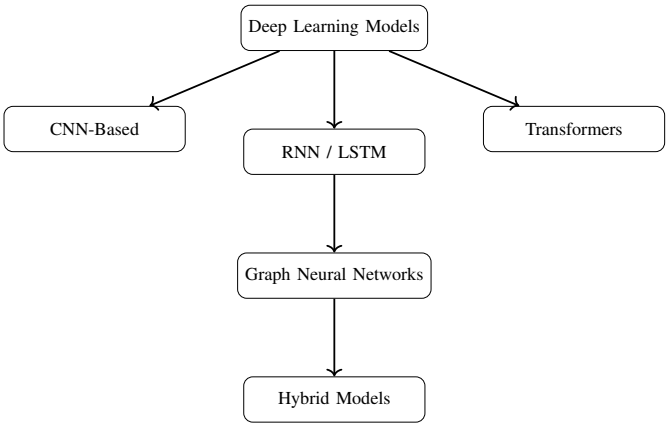

Fig. 3. Taxonomy of Deep Learning Models for Dark Web CTI

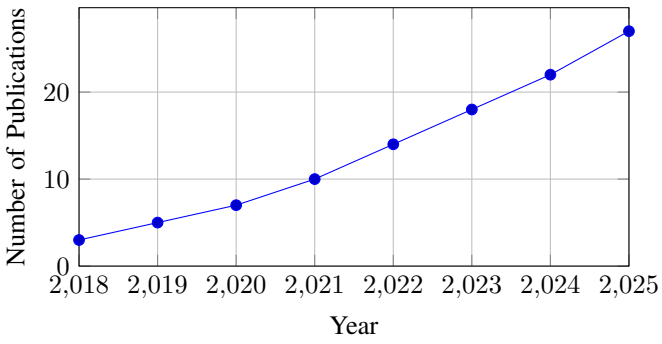

Fig. 4. Research Trend in Deep Learning–Based Dark Web Studies (2018–2025)

tion across different darknet settings [16], [43], [44].Despite these constraints, deep learning regularly performs better than traditional methods when paired with information-rich datasets such as CIC-Darknet2020 and UTD-Darknet [45], [46].

Hybrid architectures have been investigated to address deep learning weaknesses. Combining statistical features with neural models enhances generalization and resistance to adversarial darknet activities [47].Federated learning methods maintain privacy while supporting distributed darknet threat detection [48].

Overall, existing research indicates a clear move toward end-to-end deep learning pipelines that include attention mechanisms, multimodal representations, and real-time threat detection abilities. These developments serve as the basis of the proposed deep-learning-driven cyber threat intelligence platform.

Deep Learning Breakthroughs in Dark Web Intelligence (2025) suggests a system for automated data extraction from darknet markets using fine-tuned NER models (ELMo-BiLSTM, UniversalNER, GLiNER), reaching about 91% precision, 96% recall, and 94% F1-score on a new annotated DNM dataset, to support law enforcement in tracking illicit trade. [49]The article DeepWebNet Analytics for Dark Web Surveillance (2025) presents a pipeline that combines transformer-based NLP preprocessing with Attention-based Graph Convolutional Networks and anomaly detection to process unstructured, partially encrypted dark web data to monitor malicious activity almost in real-time. [49]The article DarkMor Framework for Darknet Traffic Detection (2024) suggests a deep learning model that integrates local and spatial network features, achieving about 97.78% accuracy on the CIC-Darknet2020 dataset for darknet traffic classification and improving robustness under anonymity and encryption constraints. [50]DarknetSec: Self-Attentive Deep Learning (2022) introduces a fusion of 1D CNN, Bi-LSTM, and self-attention for darknet traffic detection, showing superior F1-scores over baseline models on encrypted and difficult dark web traffic patterns. [51] Scraping the Shadows: this articlr NER on Dark Web (2024) examines fine-tuned NER models on darknet market datasets for deriving entities such as drugs, prices, and vendors, reporting up to roughly 94% F1-score and emphasizing automated intelligence gathering over manual scraping. [52] Darknet Traffic with Stacking Ensembles (2023) designs a altered deep learning stacking ensemble to analyze and classify darknet network flows, surpassing 97% accuracy for threat detection and effectively distinguishing dark web activities from benign traffic. [53] DarkNET Traffic Detection Using Deep Learning (2025) combines deep learning classifiers with Bayes estimators to improve the reliability of DarkNET traffic verification across different network architectures, enhancing detection performance in high-anonymity environments. [54] Dark Web Activity Classification Using Deep Learning (2023) applies deep neural networks and CNN algorithms for image recognition on extracted dark web images, enabling classification of illicit activities and improving automated threat identification through visual pattern analysis. [55]

III. Challanges and Limitations

Research on the dark web and the development of automated Cyber Threat Intelligence (CTI) systems face many challenges that make them very different from traditional web analysis. One major problem is limited and restricted data access, as dark web content is only reachable through anonymizing networks like Tor and is surrounded by ethical, legal, and privacy concerns. Strict regulations often limit continuous data collection. In addition, dark web platforms are highly unstable—they frequently shut down, reappear under new names, or change addresses. This constant change makes previously collected data quickly outdated and causes machine learning models to lose accuracy over time. Another difficulty is the lack of labeled data, since manual labeling is time-consuming and exposes researchers to illegal or harmful content. As a result, many datasets are small or poorly labeled, which reduces the effectiveness of supervised learning models. Technical challenges further complicate CTI system development. Dark web users rely heavily on encryption, anonymization, and disguised language, making feature extraction and analysis difficult and computationally expensive. Dark web intelligence is also multimodal, involving text, images, transactions, and network behavior, which requires complex and resource-intensive systems to combine different data types. Finally, legal

and regulatory limitations strongly affect what data can be collected and shared, especially across borders, forcing CTI systems to carefully balance intelligence gathering with legal compliance.

IV. Conclusion and Future Work

The dark web, which facilitates the exchange of illicit commodities, malicious services, and stolen data, is still a significant source of cyberthreats. The increasing sophistication of cybercriminals makes it difficult for traditional security tools to keep up.In order to determine what has been achieved and where there are still gaps, this review compiled the literature on deep learning techniques, machine learning approaches, and dark web datasets. The investigation reveals that although several helpful techniques for categorizing dark web content and identifying suspicious activity have been created by researchers, the majority of current solutions still fall short of what is required for practical, operational cyber threat intelligence.The models that are currently used do not provide visibility and clarity to the analysts. For this the detection systems must take some certain measures like the systems must be flexible, scalable, and updated on a regular basis due to the nature of the darkweb . As a result the study Focuses on the need dark web intelligence systems that must be new as well, ones that provides a clear explanantions, combine many data sources and are designed to work in real-time environments. In future the systems can provide more accurate detection of the threat, better enhance the capabilities of cybersecurity teams, and provide a safer digital network by resolving these challenges and research gaps.

References

[1] M. Husák, J. Komárková, E. Bou-Harb, and P. Čeleda, "Survey of attack projection, prediction, and forecasting in cyber security," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, 2019.

[2] M. Juarez, S. Afroz, G. Acar, C. Diaz, and R. Greenstadt, "A critical evaluation of website fingerprinting attacks," in *Proceedings of the ACM Conference on Computer and Communications Security (CCS)*, 2014.

[3] A. ElBahrawy and et al., "Collective dynamics of dark web marketplaces," *arXiv preprint*, 2019.

[4] S. Samtani, R. Chinn, and H. Chen, "Detecting illicit content on dark web forums using transformer models," *IEEE Access*, vol. 9, 2021.

[5] S. Wegberg and R. Verburgh, "Dark web marketplaces and covid-19: before the vaccine," *EPJ Data Science*, 2021.

[6] D. McCoy and et al., "Shining light in dark places: Understanding the tor network," in *Privacy Enhancing Technologies Symposium*, 2008.

[7] J. Kim and et al., "Developing a dark web collection and infrastructure for cyber threat intelligence," Naval Postgraduate School, Tech. Rep., 2020.

[8] A. Greenberg, "Operation bayonet: Inside the sting that hijacked an entire dark web drug market," WIRED, 2017.

[9] P. Koshy, D. Koshy, and S. Capkun, "Deanonymizing tor hidden service users through bitcoin transactions," *ScienceDirect*, 2018.

[10] J. Z. et al., "Odtc: Online darknet traffic classification using multimodal self-attention," 2023.

[11] J. Aldridge, *The Evidentiary Trail Down Silk Road*. Springer, 2017.

[12] L. Baravalle *et al.*, "Developing a dark web collection and infrastructure," SciSpace Technical Report, Tech. Rep., 2020.

[13] Y. Zhang, J. Aldridge, and D. Decary-Hetu, "Dark web markets: Turning the lights on alphabay," in *Proceedings of the ACM Digital Library*, 2018.

[14] R. Tamboli and N. Bendale, "Vaccines and more: The response of dark web marketplaces to external events," *PLOS ONE*, vol. 17, no. 4, p. e0266000, 2022.

[15] A. Greenberg, "The cops laid a trap and alphabay users waltzed right into it," WIRED Magazine, 2017.

[16] N. A. et al., "A comprehensive survey on darknet analysis, challenges, and machine learning approaches," 2022.

[17] A. Khan, A. Gani, and A. W. A. Wahid, "Towards safe cyber practices: Developing a proactive cyber-threat intelligence system for dark web forum content," *Computers & Security*, 2021.

[18] A. Manolache, F. Brad, A. Barbalau, R. T. Ionescu, and M. Popescu, "Veridark: A large-scale benchmark for authorship verification on the dark web," in *NeurIPS Workshops*, 2022.

[19] S. Samtani, R. Chinn, and H. Chen, "Shedding new light on the language of the dark web (coda dataset)," in *ICWSM*, 2022.

[20] D. Décary-Hétu and J. Aldridge, "A large-scale longitudinal structured dataset of the dark web cryptomarket 'evolution'," *Journal of Cybersecurity*, 2015.

[21] J. Bakermans, P. D. Pascale, J. Marcelino, G. Cascavilla, and Z. Geradts, "Scraping the shadows: Deep learning breakthroughs in dark web intelligence," *arXiv preprint*, 2025.

[22] A. Yazdanjue, T. Zhang, S. Salinas, J. Ochoa, Y. Zhou, V. Nguyen, and D. J. Song, "A language model-driven semi-supervised ensemble framework for illicit market detection," *arXiv preprint*, 2025.

[23] G. R. Group, "Duta-10k dataset: Dark web urls classification dataset," 2022.

[24] C. Lab, "Cic-darknet2020: Darknet traffic classification dataset," 2022.

[25] CAIDA, "Ucsd/caida network telescope and darknet scanning dataset," 2022.

[26] S. Consortium, "Safesurf darknet 2025 dataset: Tor, i2p, freenet traffic," 2025.

[27] D. B. et al., "Trends over a decade of the dark web drug trade," *Journal of Drug Issues*, 2023.

[28] M. V. et al., "Cheaper than you thought? a dive into the darkweb market," in *ARES Conference*, 2023.

[29] T. M. et al., *Hydra: Lessons from the World's Largest Darknet Market*. Wiley Online Cybercrime Studies, 2023.

[30] A. S. Kumar, "White house market data collection and analysis," 2023.

[31] N.-N. P. Team, "Nd-ner: A named entity recognition dataset for osint," 2023.

[32] L. Santoso and K. Patel, "Threat intelligence named entity recognition: A comprehensive survey," *SpringerLink*, 2024.

[33] P. W. Team, "Authorship verification and identification benchmark datasets," in *Webis*, 2023.

[34] H. Zhang and M. Sap, "Who wrote when? author diarization in social media," in *ACL Anthology*, 2024.

[35] NDARC, "Cryptomarket data collection: Methods and dataset notes," UNSW Technical Report, Tech. Rep., 2023.

[36] S. K. R. et al., "Darkweb research: Past, present, and future trends," *ScienceDirect*, 2023.

[37] M. S. et al., "Darkdetect: Darknet traffic detection and categorization using modified cnn-lstm," 2021.

[38] A. B. et al., "Deep hybrid cnn-lstm for tor traffic classification," 2021.

[39] H. C. et al., "Contextualgraph-llm: Multi-modal darknet threat analysis using graph neural networks and transformers," 2024.

[40] R. K. Shalini and P. Kumar, "Bert-based intent detection for illicit dark-web forum posts," 2022.

[41] F. G. et al., "Gan-generated darknet traffic for enhanced model training," 2023.

[42] S. Ahmad and H. Malik, "Autoencoder-based anomaly detection in tor network traffic," 2021.

[43] U. S. et al., "Darknet traffic analysis: A systematic literature review," 2023.

[44] S. S. et al., "Machine learning-powered darknet threat detection: A survey," 2022.

[45] M. S. et al., "Cic-darknet2020 dataset for darknet traffic classification," 2021.

[46] U. C. Lab, "Utd-darknet dataset release for encrypted traffic analysis," 2022.

[47] R. R. et al., "Hybrid statistical–deep neural approach for darknet traffic classification," 2023.

[48] P. S. A. et al., "Federated deep learning for privacy-preserving darknet detection," 2024.

[49] S. Kumaran *et al.*, "Deepwebnet analytics enhancing dark web surveillance for effective cyber threat detection and mitigation," *International Journal of Digital Earth*, vol. 18, no. 1, p. 2495033, 2025, transformer NLP + AGCN for anomaly detection [web:1].

[50] Anonymous, “Darkmor: A deep learning framework for darknet traffic detection,” *Neurocomputing*, 2024, 97.78
[51] ——, “Darknetsec: Self-attentive deep learning model for darknet traffic detection,” *Computers & Security*, 2022, cNN-BiLSTM-self-attention outperforms baselines [web:3].
[52] ——, “Scraping the shadows: Deep learning breakthroughs in dark web intelligence,” *SSRN Electronic Journal*, 2024, nER on DNM datasets up to 94
[53] ——, “Darknet traffic analysis and classification system based on modified stacking ensemble learning,” *Heliyon*, vol. 9, no. 2, p. e13492, 2023.
[54] ——, “Darknet traffic detection using deep learning,” *IEEE Conference Proceedings*, 2025.
[55] ——, “Dark web activity classification using deep learning,” *arXiv preprint arXiv:2306.07980*, 2023.