Journal of Frontiers in Multidisciplinary Research

The Role of Network Analytics and Dark-Web Surveillance in Identifying Potential Victims, Tracking Human-Trafficking Trends, and Disrupting Organized Trafficking Networks

Ogechi Abah
College of Business and Management, University of Illinois Springfield 62703, Illinois, United States

* Corresponding Author: **Ogechi Abah**

Article Info

E-ISSN: 3050-9726
P-ISSN: 3050-9718
Volume: 07
Issue: 02
Received: 30-05-2026
Accepted: 29-06-2026
Published: 28-07-2026
Page No: 137-154

Abstract

Background: In the last two decades, human trafficking has increasingly taken place on digital infrastructure, with recruitment, advertisement, coordination, and payment increasingly being done by way of social media, encrypted messaging apps, open-web classified sites, and dark-web forums. The field of investigative practice has adapted to these two lines of analysis, namely social network analysis (SNA) and dark-web/open-source digital surveillance, both of which have an evidence base that has yet to be synthesized together with a set of research questions.
Objective: This study aims to identify potential trafficking victims, trace the evolution of trafficking patterns, disrupt organized networks of trafficking and exploitation using network analytics and dark-web surveillance, and describe the ethical and privacy protections needed for responsible use of these tools.
Methods: A qualitative, document-based research design was used. Systematic searching and screening of peer-reviewed journal articles, preprints, and institutional reports (United Nations Office on Drugs and Crime, International Labour Organisation, European Commission, Eurostat) and reputable practitioner sources (Chainalysis, Elliptic, Polaris Project) were conducted, followed by an analysis of the sources in relation to five pre-specified research questions. The primary surveys, interviews and original network datasets were not collected, and the unit of analysis was the published study or statistical release.
Results: It was found that certain structurally important actors whose removal would provide disproportionately disruptive value to the network could be repeatedly identified using centrality-based network methods and selectively enforced. Construct validity and label reliability were questioned, but increasingly powerful machine-learning classifiers were based on online advertisement textual, visual, and stylometric features paired with their relational features. Blockchain analytics introduced a brand-new evidentiary channel – the cryptocurrency ledger is not only permanent but also public, even if the wallet owner is not. Platform-policy interventions that tried to curb trafficking advertising, primarily the closure of Backpage and the passage of FOSTA-SESTA, have had mixed and at times counterproductive results, both in terms of victim identification and in terms of providing safety to the general population who engage in sex work. All the analytic gains found were paired with proven risk of algorithmic bias, privacy violations, and misidentification.
Conclusion: Summary: Network analytics and dark-web surveillance are more properly viewed not as automated systems of surveillance, but as intelligence-augmentation systems that must be monitored, authorized, and protected, survivor-by-survivor. This study aims to address the following 5 research questions sequentially in Sections 4 through 9.

DOI: https://doi.org/10.54660/.JFMR.2026.7.2.137-154

Keywords: human trafficking, social network analysis, dark web, open-source intelligence, blockchain forensics, victim identification, organized crime disruption, centrality, natural language processing, platform policy

1. Introduction
1.1. Background and Scope of Human Trafficking as a Global Crisis
1.1.1. Defining Trafficking Across Its Recognized Forms
According to the United Nations Convention against Transnational Organized Crime, trafficking involves recruiting, transporting, transferring, harboring, or receiving individuals by using force, fraud, or coercion for exploitation purposes.

This exploitation can manifest in numerous ways, such as commercial sexual exploitation, forced labor, domestic servitude, forced begging, forced marriage, organ trafficking, and coerced participation in criminal activities, affecting both adults and children (Zimmerman & Kiss, 2017) [55]. This definitional breadth is analytically significant, as each form of exploitation leaves behind a very different digital and relational footprint: Forced criminality is rarely advertised, sexual exploitation is more likely to be listed in classified and adult-service platforms, and forced labor is more likely to be found in records of recruitment agencies and patterns of wage remittance. In light of this, it is important to note that a review of network analytics and dark-web surveillance should be approached with the understanding that it is most applicable to forms of trafficking that produce lots of digital trails, but only partially applicable to those that take place more often in a space that is largely not visible online.

1.1.2. Scale and Recent Trajectory of the Crisis

Human trafficking is a large and growing global challenge. In 2021, an estimated 49.6 million people were in forced labor and 22 million in forced marriage on any given day, up from 37.9 million in 2016 (International Labour Organization *et al*., 2022) [24]. Of all forced labor, 23% involved forced commercial sexual exploitation, with nearly 80% of those victims being women or girls (ILO, 2022a).

Data on detected cases also show increases. The UNODC's 2024 Global Report, drawing on information from 156 countries and over 1,000 court cases, found a 31% rise in child trafficking victims between 2019 and 2022 (United Nations Office on Drugs and Crime, 2024) [45]. In the EU, registered victims rose from 7,155 in 2021 to 10,093 in 2022 (+41%), then rose further to 10,793 in 2023 (+6.9%), the highest total recorded in the 2008–2023 period (Eurostat, 2024; European Commission, 2025b) [19, 17].

Table 1 summarizes the forms of exploitation covered in this review, with selected definitions

Table 1: Forms of Exploitation Addressed in This Review

Form of exploitation	Working definition	Representative source
Commercial sexual exploitation	Exploitation through the sale of sexual services obtained by force, fraud, or coercion, or involving a minor regardless of consent	Polaris Project (n.d.)
Forced labor	Work or service exacted under threat of penalty, without voluntary offer	International Labour Organization *et al*. (2022)
Domestic servitude	Forced labor within a private household, typically involving isolation and restricted movement	Polaris Project (2025)
Forced begging	Coerced solicitation of money or goods from the public for a third party's benefit	United Nations Office on Drugs and Crime (2024)
Forced marriage	Marriage without free and full consent of one or both parties, often accompanied by exploitation	International Labour Organization *et al*. (2022)
Organ removal	Trafficking to extract organs or tissue without informed, voluntary consent	United Nations Office on Drugs and Crime (2024)
Coerced criminal activity	Compelling a trafficked person to commit crimes (e.g., cultivation, fraud, theft) for exploiters	European Commission (2024)

1.2. The Digital Transformation of Trafficking Operations

1.2.1. Platform Migration and Recruitment Channels

Trafficking investigations are complex and are unique compared to many types of organized crime, such as evidence is spread across jurisdictions, communication being often anonymous or encrypted, identities are often fabricated or stolen, and ads are rotated frequently to avoid takedown, as well as the payment often being transacted in digital or cryptocurrency (Antonopoulos *et al*., 2020) [1]. Traffickers have quickly embraced the platforms available, as over two-thirds of underage victims recruited online in 2020 were recruited via Facebook, 14 percent via Instagram, and eight percent via Snapchat (as described in the darknet and criminal-justice literature consulted below). Such a platform migration is not a one-way street, however, and as each platform attempts to reduce its moderation or is hit by legal action, recruitment and advertising activity has been seen to move to newer or less-moderated platforms repeatedly, a migration that will be explored in more depth in Section 5.5.

1.2.2. The Dark Web as an Additional Operational Layer

Dark-web forums, marketplaces, and messaging services that are accessible via anonymizing software like Tor have formed an extra layer of the ecosystem, where advertisements are posted and buyers coordinated, as well as increasingly, to launder profits generated from exploitation via cryptocurrency (Medipelly & Abosata, 2024) [29]. This is a liability as well as an asset when it comes to investigators and an opportunity. The same data that is used by traffickers to grow their operation also fuels the raw data that can be used in network analysis and compliance monitoring: phone numbers used in many ads, photos reused in many ads, cryptocurrency addresses that conduct transactions with known illicit services. This duality inspires the present review. Recent research highlights the tension between centralized data aggregation and the need for agile, decentralized AI architectures, suggesting that hybrid approaches combining data locality with distributed model deployment may be particularly valuable for complex data-intensive applications (Tasleem *et al*., 2025) [42].

1.3. Purpose and Structure of This Study

This study examines two connected analytic approaches:

(a) social network analysis applied to trafficking operations, and

(b) dark-web and open-source digital surveillance, including natural-language processing of advertisements and blockchain forensics of payments.

Section 2 introduces the conceptual and theoretical foundations, including the widely used Zimmerman *et al*. (2011) trafficking-process model and standard graph-theoretic centrality measures.

Section 3 describes the research design and methodology.

Sections 4 through 8 present findings organized around the five research questions specified in Section 1.5.

Section 9 discusses the findings and their implications. Sections 10 and 11 address limitations, recommendations, and conclusions.

1.4. Problem Statement

Although significant advances have been made in the techniques of social network analysis, natural-language processing, image analysis, and blockchain forensics over the last decade, no single study has integrated all four analytic areas and assessed them against a unified set of investigative goals: victim identification, trend detection, and network disruption. Existing research tends to be fairly narrow and focused, e.g., a computer-science paper may focus on the accuracy of a single classifier, a criminology paper may focus on reconstructing a single network, a financial-crime report may focus on the laundering cluster of a particular financial institution, etc., without providing a unified view of what these techniques can and cannot collectively achieve, or without discussing the ethical and evidentiary considerations that come into play when using them together in practice. This study aims to fill that gap.

1.5. Research Questions

This study was guided by the following five research questions:

RQ1: How can network analytics be used to identify structurally important actors and disrupt organized human-trafficking networks?

RQ2: How can dark-web surveillance, natural-language processing, image analysis, and blockchain forensics be used to detect trafficking-related advertisements and financial flows?

RQ3: What behavioural, contextual, and digital indicators support the recognition of potential trafficking victims and situations?

RQ4: What do global and regional data reveal about recent trends in detected human trafficking?

RQ5: What ethical, privacy, and bias risks arise from network-analytic and dark-web surveillance methods, and what safeguards mitigate them?

1.6. Research Objectives

Based on the above research questions, this study had five objectives:

1. Review and compare the effectiveness of centrality-based and community-detection network methods in identifying key actors and disrupting trafficking networks;
2. Review and compare the capacity of machine-learning, natural-language-processing, image-analysis, and blockchain-forensic methods to identify trafficking-related ads and financial flows;
3. Consolidate the behavioural and digital indicators documented in the literature for the identification of potential victims and the identification of trafficking situations;
4. Characterize the recent trends in identified trafficking globally and regionally using the best available primary statistical sources; and
5. Identify the ethical, privacy, and bias risks documented in the literature, along with proposed safeguards.

1.7. Significance of the Study

This study is significant for three audiences. It brings together a disparate and inter-disciplinary literature from criminology, computer science, and public health and explains the implications of the methodological insights that can be applied across disciplines and those that cannot. It offers a framework and source-referenced summary of tools and techniques and their reported performance, as well as a straightforward summary of their documented limitations, to inform realistic expectations when these tools are used in practice. It brings a stark and actionable dilemma to the fore for policymakers: the tension between the visibility platforms are a prerequisite to detecting trafficking ads, and their suppression at the platform level is essential to the fight against trafficking. This creates a unique and practical paradox for designers of anti-trafficking laws and platform liability rules in the future: visibility is a prerequisite to detection, but suppression at the platform level is necessary to combat trafficking.

2. Literature Review and Theoretical Foundations

2.1. The Trafficking Process Model and the Lived Experience of Exploitation

2.1.1. Stages of the Zimmerman *et al*. Framework

One of the most often mentioned conceptual models was originally proposed by Zimmerman *et al*. (2011) [56] to link trafficking with health outcomes and has been subsequently adapted in the field; it considers trafficking as a multi-stage process, not just a single event. The stages are typically depicted as pre-trafficking, recruitment/travel/transit, exploitation, and life post-trafficking, and can be re-trafficking or re-integration (as adapted from Zimmerman *et al*., 2011; Klabbers *et al*., 2023) [56, 28]. This structure is reproduced in Fig. 1.

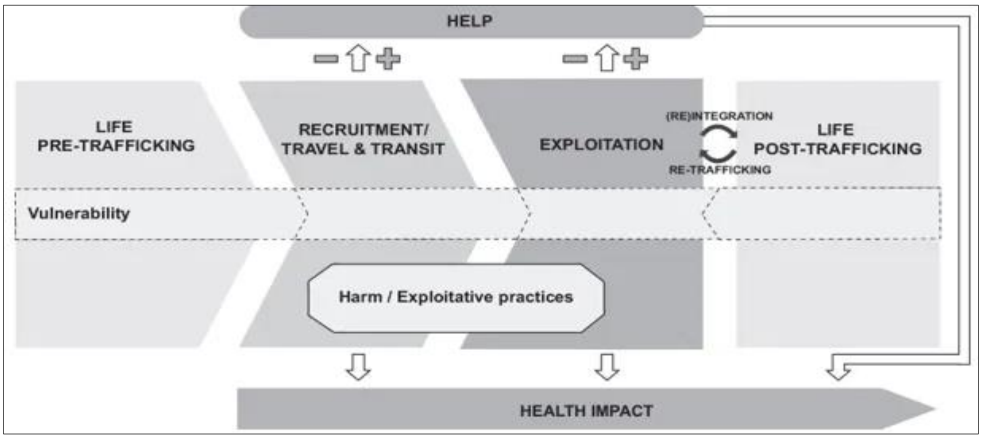

Fig 1: Human trafficking process framework, adapted from Zimmerman *et al*. (2011) and reproduced from a subsequent health-focused adaptation (Klabbers *et al*., 2023).

2.1.2. Vulnerability, Harm, and Health-Impact Pathways

The importance of this stage model for network analytics is that it can help to determine where digital traces are likely to be created and where intervention is most likely to be possible. Vulnerability builds up before trafficking, recruitment, travel, and transit create communications and financial traces, exploitation creates advertisements, controlling communications and payments, and opportunities for referral and, unfortunately, re-trafficking risk are created during the post-trafficking stage. The framework in figure 2 is extended with the specific vulnerability factors, control tactics and health impacts documented in the lived-experience literature focused on forced labor and sex trafficking victims.

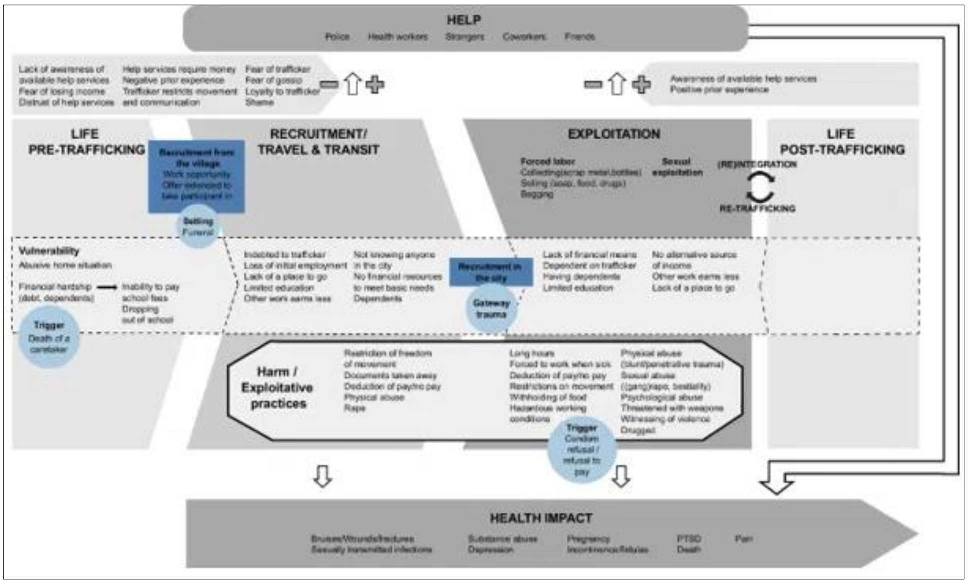

Fig 2: Patterns in the lived experience of forced labor and sex trafficking victims, mapped onto the trafficking-process framework.

2.2. Social Network Analysis as an Analytic Framework

2.2.1. Historical and Disciplinary Origins of Social Network Analysis

Social network analysis (SNA) is a framework that sees a criminal enterprise as a graph, with the nodes as people, phone numbers, online accounts, cars, houses, ads, etc., and the edges as documented relations or interactions between the nodes (Cockbain *et al*., 2011) [8]. Drawing on the thinking of sociology, the approach is broadly expanded, in particular to the study of organized crime in general and human trafficking in particular, on the premise that people are situated within 'thick webs of social relations and interactions' that shape their behaviour and vulnerability to being disrupted (Borgatti *et al*., 2009 as cited in Cockbain *et al*., 2011). A recent thesis-length synthesis found that SNA is useful for disrupting trafficking networks, specifically by allowing investigators to target structurally important figures, as well as the fact that there are ongoing gaps in the research on the structure of online trafficking networks, rather than offline criminal groups in general (Vanderbilt Institutional Repository, 2022) [52].

2.2.2. Static Versus Dynamic Network Perspectives

A further conceptual distinction, developed in the organized-crime literature more broadly, separates static network description from dynamic, disruption-oriented analysis. A computational-modelling study combining social network analysis with criminal-intelligence data from the Dutch police found that, contrary to common assumption, targeted removal of key actors can leave criminal networks just as functional, or even more efficient, afterward, underscoring the need to consider network dynamics and how networks shift and bounce back when faced with law enforcement action (Duijn, Kashirin, & Sloot, 2014) [10]. This is of direct relevance to trafficking, where networks can be seen to reform after the arrest of some of their members, as is described in 4.3.2.

3. Research Design and Methodology

3.1. Research Design

The research design for this study was a qualitative, document-based research design, with the published study, institutional report, or statistical release being the unit of analysis as opposed to the individual, organization, or original data collected for the study. The design was chosen because the five research questions outlined in Section 1.5 are not questions that would be best addressed via new primary data collection (survey, interviews, or original network reconstruction), but rather questions about the state of the consolidated evidence base (spanning criminology, computer science, and public health). A survey, an interview, or a primary network-analytical dataset were not conducted for this study; all quantitative statistics cited in Sections 4-8 are cited directly from the cited primary sources.

3.2. Data Sources and Search Strategy

Structured, iterative research efforts of peer-reviewed journal articles, preprint repositories (mostly arXiv), institutional reports (United Nations Office on Drugs and Crime, International Labour Organization, European Commission/Eurostat), and reputable practitioner and industry sources (Chainalysis, Elliptic, Polaris Project, ShadowDragon) were used to identify data sources. The research questions were matched to the search terms that included the core concepts of human trafficking, social network analysis, centrality, dark web, open-source intelligence, cryptocurrency forensics, platform policy, and victim identification.

3.3. Inclusion and Exclusion Criteria

Sources were selected when they were able to describe empirical applications of the network analytic or digital-surveillance approaches to trafficking, when they provided original prevalence or trend statistics from an identifiable primary data collection (e.g. Eurostat, the UNODC Global Report series, or the ILO–Walk Free–IOM global estimates program), or when they provided conceptual or ethical framing directly relevant to one of the five research questions. Grey literature and vendor material were selectively included, were restricted to material that included an identified investigative method or documented case outcomes, and not weighted as highly as peer-reviewed sources in the synthesis presented in Sections 4 – 8. Sources with no known author, institution, or organization, or which made claims that one could not verify in a second source, were omitted.

3.4. Data Analysis Procedure

Retained sources were read through in full and findings relevant to each of the research questions were extracted and presented thematically under each research question, as outlined in Sections 4-8. When more than one source provided the same statistic, the quantitative reporting results were cross-checked and confirmed with another source before being included (e.g., the European Union victim-count figures in Section 7.3 were checked and confirmed against two independent releases from Eurostat and the European Commission before being included). This is explicitly noted where a single source has not been substantiated, and is not included as a consensus finding. Findings were synthesized across studies using thematic synthesis, and not statistical meta-analysis, given the variety of methods and units of measurement across the underlying literature.

4. Results: Network Analytics in Trafficking Investigations (RQ1)

4.1 Identifying Central Actors and Key Nodes

The dominant use case for SNA in trafficking investigations is the identification of structurally important actors as targets for further investigation, arrest, or prosecution.

4.1.1. UK Internal Child Sex Trafficking Case Findings

Cockbain *et al*. (2011) [8] analysed and utilized centrality measures for the purposes of helping investigators target child sex offenders and to balance their assumptions or biases based on previous experience with a child victim independent of network structure, in two cases of internal child sex trafficking in the United Kingdom, where they collected data on 25 offenders and 36 victims during police investigations.

4.1.2. Latino Sex Trafficking Enterprise Case Findings

A subsequent study similarly addressed related questions, but in a different context: A SNA study conducted by Sabon, Yang and Zhang (2021) in a regional Latino sex trafficking enterprise explored what network structures were associated with collaboration between traffickers and victims, and found that the network structures were not consistent with how this intra-ethnic immigrant network was organized, which further suggests that network structure in sex trafficking operations is not uniform across cultural or regional contexts (see Vanderbilt Institutional Repository, 2022) [52].

4.1.3. Lessons from Organized-Crime Network Studies More Broadly

In addition to the specific area of sex trafficking, SNA has a rich history in the study of organized crime in general, particularly in Mafia-like organizations, providing methodological lessons relevant to the analysis of trafficking networks. In an experiment, Ficara, Curreri, Fiumara, and De Meo (2022) [21] compared interventions aimed at actors with high social capital, interventions aimed at actors with high human capital (specialized criminal skill or experience), and random node removal to seven different disruption strategies across two real Mafia networks reconstructed from judicial records. Their key finding was that the effectiveness of targeted removal — either by social or human capital (key actors) — was greater than random removal in reducing network cohesion, but that network fragmentation differed when either social or human capital targeting was used, calling for the disruption strategy to be suited to the type of network weakness that the investigator wished to exploit, rather than being a standardized formula.

4.2. Community Detection and Subgroup Structures

4.2.1. Cellular Structure of Trafficking Operations

Trafficking networks are not monolithic; they tend to be composed of sub-groups of trafficking in which each sub-group consists of a smaller number of actors, such as recruitment cells, transportation cells, exploitation sites, and financial layers, who connect through a smaller number of bridging actors. The mobility data detection literature summarizes the recovery of this cellular structure from raw relational data, for example, shared phone numbers, shared images, co-occurrence in the same ads, etc., using community detection algorithms (Keskin *et al*., 2021). Investigatively, this structural decomposition can be useful to uncover the new relationships between cases that were seemingly unrelated to each other.

4.2.2. Advertisement Clustering as a Proxy for Network Structure

When relational data (who communicated to whom) is not available, the researchers have used features at the level of advertisements as a proxy for the structure of the network. Keskin *et al*. (2021) categorized ads by common text, phone numbers, and image hashes, assuming that ads with similar text, phone numbers and image hashes are essentially from the same underlying operator or cell, and then used the resulting categories to forecast the next location of ads based on frequency-based analysis and network simulation. In this way, text-classification methods (Section 5.2) and network-analytic methods are not competing techniques, but are complementary in practice.

4.3. Network Disruption Strategies and Their Comparative Effectiveness

A recurring finding across the organized-crime SNA literature is that networks are adaptive: removing a central actor does not guarantee lasting disruption if the network can promote or recruit a replacement into the vacated structural position.

4.3.1. Social-Capital and Human-Capital Targeting

Ficara *et al*. (2022) [21] treated targeting strategies in two categories: targeting based on an actor's social capital (the density and reach of their personal networks) and targeting based on an actor's human capital (specialized criminal skill, experience, or technical knowledge that can be difficult for the network to replace). Their comparative testing results across two Mafia networks revealed that both sets of targeted removal are more successful than random removal, but that there are identifiable fragmentation signatures for each type of capital, suggesting that investigators with intelligence about which type of capital is dominant in a particular Mafia network may choose a more finely targeted approach to removal. The adaptive recovery is a concept that hangs on the limits of the one-time disruption.

4.3.2. Adaptive Recovery and the Limits of One-Time Disruption

Duijn, Kashirin, and Sloot (2014) [10] explicitly called for more focus on the adaptive capacity of networks, suggesting that disruption strategies should not be judged solely based on their impact on network cohesion but on how well the network recovers over time. As in organized-crime research in general, this concern is one of the potential shortcomings of a single well-structured high-profile arrest; if there are "replacement actors" available to quickly replace the high-profile actor caught, then the impact of the arrest may be only temporary, without actual dismantlement, and longitudinal, repeated-measurement designs are required to separate such temporary disruption from true dismantlement.

4.3.3. Financial-Coordinator Targeting in Conflict-Affected Trafficking

There is an analogous concern in trafficking-specific settings: a 2025 data-driven study on mapping trafficking networks that arose around the Russia–Ukraine conflict suggested that the destabilisation of central financial coordinators, as opposed to peripheral ones, would be more likely to lead to network-wide destabilisation, using a combined open-source intelligence, social-network-analysis, and blockchain-analysis methodology (arXiv, 2025). Table 2 shows the disruption strategy findings from the studies reviewed.

Table 2: Comparative Findings on Network Disruption Strategies

Study/network context	Disruption strategies compared	Key reported finding
Ficara *et al*. (2022) — two reconstructed Mafia networks	Social-capital targeting; human-capital targeting; random node removal	Both capital-based targeted strategies outperformed random removal, but produced different fragmentation patterns
Duijn *et al*. (2014) — Dutch police criminal networks, general framework	Static centrality-based removal vs. dynamics-aware, adaptation-informed disruption	Argued static approaches under-weight network recovery and called for dynamic, longitudinal disruption models
Cockbain *et al*. (2011) — UK internal child sex trafficking	Centrality-guided investigative targeting vs. experience-based targeting	Centrality metrics identified key actors and reduced reliance on investigator assumptions
Mapping Trafficking Networks (2025) — post-conflict trafficking, OSINT/SNA/blockchain	Targeting central financial coordinators vs. peripheral actors	Destabilizing central coordinators was associated with broader network instability

4.4. Illustrative Case Applications

4.4.1. Divergence Between Regional Case Studies

The reviewed case studies are examples of what network-analytic disruption can do and where it can stop. Cockbain *et al*. (2011) [8] found that the properties of the offender and victim networks were different in the UK internal trafficking cases they studied, suggesting that there could be no single disruption template that could be applied to both sub-networks. Collaborations among traffickers and victims were found to be different for the Latino sex trafficking enterprise studied by Sabon *et al*. (2021) because they were shaped by shared ethnicity and immigration status which yielded other centrality distributions than found for the UK cases reviewed; this difference was explicitly called out by later reviewers as contradicting an assumption of structural uniformity that had been drawn from the earlier study (Vanderbilt Institutional Repository, 2022) [52].

4.4.2. Implications for Generalizability

Taken together, the cases reviewed here conclude that there is a central methodological takeaway from this review: a network-analytical result from one cultural context, region, or exploitation-type does not necessarily hold in another. Investigators and researchers using centrality-based targeting techniques in a novel context should consider previous results to be hypotheses that can be tested and validated with the local relational data collected.

5. Results: Dark-Web Surveillance and Digital Evidence (RQ2)

5.1. Nature of the Dark Web and Its Use by Trafficking Networks

5.1.1. Anonymizing Infrastructure and Market Scale

The dark web is a place where people post content using darknets and can access it by using specialized anonymizing software like Tor, I2P, or Riffle, and it can be used for the advertisement, coordination, and payment of trafficking, in addition to drug and weapons trading and other illegal activities (Medipelly & Abosata, 2024) [29]. In 2022, the total value of the darknet market was estimated at approximately 1.5 billion US dollars, and trafficking-related activity continues to thrive within this vast landscape of online commerce (Medipelly & Abosata, 2024) [29].

5.1.2. Open-Source Intelligence Beyond the Dark Web

Online marketplaces on the open web and surface web classified advertisement (CAA) sites provide an opportunity for victim identification and tracking those who participate in online facilitation of human trafficking (HRT) (ShadowDragon, 2024) [40], and online marketplaces on the open and dark web have expanded beyond the narrow confines of the dark web. This reflects the reality of how trafficking operates: the traffickers are adept at shifting their operations from the open web to mainstream social media to dark-web platforms, depending on which is expected to be least monitored at a particular time; a dark-web investigation narrowly focused on the dark web would systematically fail to capture a significant portion of relevant activity.

5.2. Machine Learning and Natural-Language-Processing

Approaches to Advertisement Analysis

A substantial body of computer-science research has applied machine learning and natural language processing (NLP) to the detection of trafficking-related advertisements, evolving over roughly a decade from simple keyword rules to sophisticated multimodal, interpretable systems.

5.2.1. Keyword and Structured-Feature Classifiers

Previous research has been based on structured features, typically using keywords to identify text that contains specific features, such as Dubrawski *et al.* (2015), which used features such as keywords and telephone numbers to train classifiers to classify advertisements as trafficking-related or not. Though this method was low cost and understandable, it was prone to quick obsolescence when traffickers simply changed their ads' text to avoid a standard list of keywords.

5.2.2. Domain-Specific Language Modeling

The first studies used structured features like keywords and phone numbers found in the text of advertisements, for instance by Dubrawski *et al.* (2015) who extracted keywords and phone numbers to train classifiers to classify between trafficking-related and non-trafficked ads. Computationally inexpensive and interpretable, but vulnerable to the rapid obsolescence of the ads because traffickers adapted their advertising language to avoid known lists of keywords.

5.2.3. Multimodal Text-Image Systems

Recent systems have included multimodal signals, with Tong *et al.* (2017) [44] proposing a multimodal deep model that included features from both text and image, and a 2023 study on classifying sex trafficking from online escort advertisements adding a multi-input deep learning model that combined text, image, and emoji features (Summers, Shallenberger, Cruz, & Fulton, 2023) [20]. Multimodal approaches are inspired by the fact that, although advertisements often use the same image, the text may be modified; image-level features can recover linkages that may be lost in text.

5.2.4. Network-Linked Advertisement Clustering

The network-analytic and text-classification approaches have also been implemented side-by-side, as presented in Section 4.2.2. Keskin *et al.* (2021) classified the text, phone numbers, and image hashes of sex advertisements in the same way as SNA, and by analysing the frequency and simulating the network, they predicted future advertisement locations.

5.2.5. Stylometric and Financial Linkage Techniques

One of the more powerful comes from a combination of authorship analysis and financial-transaction linkage. Building on this, Portnoff *et al* (2017) [37] created a stylometry classifier that could determine whether two advertisements were written by the same author or different authors, with a true positive rate (TPR) of ~90 percent at a false positive rate (FPR) of ~1 percent, and combined this with a linking technique based on data leakage from the Bitcoin mempool, Bitcoin blockchain, and the advertisement site itself to link a subset of advertisements with a particular Bitcoin public wallet and transaction. The authors showed how an analyst could apply the financial-linking technique, combined with existing hard identifiers like phone numbers, to cluster ads by true advertiser, not by the claimed advertiser as mentioned in the advertisement text, in a four-week proof of concept study on the Backpage advertising site (Portnoff *et al.*, 2017) [37]. This work stands out in the reviewed literature by explicitly linking the textual-classification tradition mentioned above to the blockchain-forensics tradition mentioned in Section 5.4.

An interpretability layer was added to advertisement classification, as part of a related, but separate, body of research. To overcome the criticism that opaque classifiers are challenging to rely on by law enforcement and courts, Rodriguez Perez and Rivas (2023) [39] proposed a natural-language-processing approach to creating pseudo-labeled training data with minimal manual supervision and an interpretability approach using Integrated Gradients for flagged advertisements to let law enforcement users inspect why the specific ad was flagged. This line of work was further extended by the development of a TrafficVis visualization system for the detection and labelling of organized activity and spatio-temporal patterns, for the purpose of supporting investigators in traffic crime detection and labelling (Vajiac *et al.*, 2022) [50], and another system, DeltaShield, which used information-theoretic anomaly detection on the same underlying advertisement data (Vajiac *et al.*, 2023) [51]. The U.S. Department of Homeland Security's Memex program, which was developed with the Defense Advanced Research Projects Agency, resulted in tools for investigators, such as TellFinder, which according to MIT News (2019) would allow investigators to index, summarize, and search sex-advertisement data, and, in at least one documented case, to identify additional underage victims from a single online advertisement, which had previously been lengthy investigations.

Table 3: Machine-Learning and NLP Approaches to Trafficking Advertisement Detection

Approach	Representative study	Core method
Keyword and structured-feature classifiers	Dubrawski *et al.* (2015)	Extracted keywords and phone numbers from ad text to train supervised classifiers
Domain-specific language modeling	Zhu *et al.* (2019)	Trained a language model on adult-service ad corpora to capture trafficking-specific linguistic structure
Context-specific BERT-based modeling	Esfahani *et al.* (2019)	Combined topic modeling, word embeddings, and BERT for automatic pattern learning
Multimodal text-image deep models	Tong *et al.* (2017)	Jointly modeled advertisement text and images for trafficking classification
Network-linked advertisement clustering	Keskin *et al.* (2021)	Grouped ads by shared text, phone numbers, and image hashes; predicted future ad locations
Stylometric authorship + financial linkage	Portnoff *et al.* (2017)	Stylometry classifier (~90% TPR, ~1% FPR) linked to Bitcoin wallets via mempool/blockchain leakage
Interpretable pseudo-labeled NLP pipeline	Rodriguez Perez & Rivas (2023)	Generated pseudo-labels with minimal supervision; used Integrated Gradients for interpretability
Investigator visualization dashboards	Vajiac *et al.* (2022, 2023) — TrafficVis, DeltaShield	Visualized organized activity and applied information-theoretic anomaly detection

5.2.6. Reliability and Construct-Validity Concerns

Alongside this technical progress, there has been technical progress, and a methodological critique published in 2024 called into question the systemic overestimation of the capacity of this field to identify trafik based on online ads alone, noting that many published studies have titles suggesting they had strong identification powers, but with annotated training labels whose inter-annotator reliability is also questionable: the same ad is sometimes annotated with different classifications by different human annotators, compromising the ground truth on which classifiers are trained and evaluated (Cogent Social Sciences, 2024) [9]. This critique does not challenge the larger analytical program, but does help to bolster a narrower conclusion: Automated systems are not analytic tools but generators of leads for human analysis, as proposed in the safeguards outlined in Section 8.

5.3. Image Analysis and Threat Detection on the Dark Web

5.3.1. Computer-Vision Classifiers for Dark-Web Content

In addition to text, there are several studies that have used computer-vision-based approaches to traffic-related images in the dark web. To use logistic regression and support-vector-machine (SVM) classifiers in an exploratory manner, Medipelly and Abosata (2024) [29] created and tested these classifiers on the dark-web-derived feature data; they concluded that while logistic regression was a strong classifier for the specific threat-classification task, SVM was even more accurate. Among the various machine-learning models evaluated for darknet application and traffic type classification, two models performed best: a LightGBM model obtained an F1 score of 93.41 percent for application type classification, and a RandomForest model obtained an F1 score of 99.8 percent for traffic type classification that outperforms other tested state-of-the-art models (ResearchGate, 2020) [38].

5.3.2. Integrated Text-Image Surveillance Architectures

Building on previous research establishing the need to simultaneously analyze text and image data in an ad to detect an ad, an Image Surveillance Assistant (ISA) architecture is proposed (dtrsociety.org, 2025) [12] to jointly analyze text and image data in an ad at scale. To summarise the quantitative performance statistics for each detection system, these are grouped together in Table 4, with the original study context included.

Table 4: Reported Quantitative Performance of Selected Detection Systems

System/study	Task	Reported performance
Portnoff *et al.* (2017) stylometry classifier	Distinguishing same- vs. different-author advertisements	~90% true-positive rate at ~1% false-positive rate
Medipelly & Abosata (2024) SVM vs. logistic regression	Dark-web threat classification	SVM outperformed logistic regression in classification accuracy
Darknet traffic/application classifiers (LightGBM; RandomForest)	Application classification; traffic classification	93.41% F1-score (application); 99.8% F1-score (traffic)

5.4. Cryptocurrency and Blockchain Financial Tracing
Because trafficking increasingly monetizes through digital assets, blockchain forensics has become an important complement to text- and image-based detection.

5.4.1. Transparency and Traceability of Blockchain Ledgers
Despite the fact that the identity of the wallet owner behind a specific wallet address is not always apparent, blockchain ledgers provide an unprecedented insight into illicit financial flows, compared to the traditional financial system, as described by Chainalysis (2022) [4].

5.4.2. Documented Laundering Clusters and Case Outcomes
From 2019 to 2022, a group of suspicious wallets sent over 13 million USD worth of Bitcoin to one another through a series of transactions and collected 29 million USD worth of Bitcoin from sales linked to the darknet marketplace Hydra, with 21 million USD of that amount then being transferred to the centralized exchanges for laundering (Chainalysis, 2022) [4]. More recently, the U.S. Treasury's Financial Crimes Enforcement Network designated Cambodia-based Huione Group, which served as a key laundering conduit for scam-compound networks that rely on human trafficking and coerced labour, as a primary money-laundering concern, finding it had laundered at least 4 billion US dollars in illicit crypto proceeds between August 2021 and January 2025, with services centered in Southeast Asia drawing victims and, in some cases, customers from across the globe (Financial Crimes Enforcement Network, 2025) [7].

5.4.3. Scam-Compound Economies and Trafficking-Adjacent Financial Crime
These financial transactions are not abstract, but they are linked to big, tangible exploitation ventures. In a second development, the U.S. Department of Justice seized some 15 billion US dollars of Bitcoin associated with a scam-compound network based in Cambodia that conducted romance-scam operations, describing it as forced labour on an industrial scale (Castellanos, 2025) [54]. This fits a broader pattern documented by INTERPOL, which found that as of March 2025 victims had been trafficked from at least 66 countries into online scam centres, with the large majority recruited into compounds concentrated in Southeast Asia (INTERPOL, 2025a) [5]. Earlier fieldwork by the UN Human Rights Office had already estimated that hundreds of thousands of people, including at least 120,000 in Myanmar and around 100,000 in Cambodia, were being held in scam-compound operations and forced to carry out online fraud (Office of the United Nations High Commissioner for Human Rights, 2023) [53].

5.4.4. Building Evidentiary Narratives from Blockchain Data
Investigative practice guides highlight that the goal of blockchain analysis is not only to trace the funds but to reconstruct a timeline of how the funds moved through time in a clear and comprehensible manner that a prosecutor and/or judge could follow, as blockchain data offers definitive proof of transfers and they are time-stamped, unlike financial records (Elliptic, 2025) [13]. The tools and platforms mentioned in the reviewed literature on the dark web and blockchain are summarized in Table 5.

Table 5: Investigative Tools and Platforms Referenced in the Reviewed Literature

Tool/platform	Analytic function	Source documenting use
Memex / TellFinder	Indexing, summarizing, and searching online sex-advertisement data for investigative leads	MIT News (2019)
Chainalysis Reactor	Clustering blockchain transactions and identifying cash-out points for laundered funds	Chainalysis (2022); INTERPOL (2025b)
Elliptic	Blockchain address tracing and chronological transaction narrative-building for court use	Elliptic (2025)
ShadowDragon Horizon / SocialNet	Open-source and social-media intelligence collection across open and dark web	ShadowDragon (2024)
CACI DarkBlue Intelligence Suite	Searchable, scraped dark-web data with secure live-access tooling for analysts	ShadowDragon (2024)
TrafficVis / DeltaShield	Visualization of organized activity and information-theoretic anomaly detection over ad data	Vajiac *et al*. (2022, 2023)

5.5. Platform Policy, Takedowns, and Displacement Effects
5.5.1. The Backpage Closure and FOSTA-SESTA
The policy landscape has evolved significantly as well, and is where the advertising-based detection and disruption strategies exist. In April 2018, the U.S. federal government shut down at the time one of the biggest classified-advertisement sites for commercial sex, Backpage, prior to the implementation of Public Law 115-164, also referred to as FOSTA-SESTA, which amended Section 230 of the Communications Decency Act (CDA) to hold online platforms liable for facilitating sex trafficking (Blunt & Wolf, 2020). As described above, the law's stated rationale was to decrease the amount of trafficking, and its intended effect was to eliminate the digital infrastructure that several of the detection capabilities studied in Section 5.2 had been designed to detect.

5.5.2. Documented Effects on Investigations and Vulnerable Populations
The outcomes of this intervention have been far from consistent and, in some ways, counter to the intended outcomes. Hacking//Hustling surveyed 98 online workers and 38 street-based workers and found that 70+ percent reported a negative financial impact from the closure of Backpage and passage of FOSTA-SESTA; 45 percent said they no longer have enough money to buy ads on the fee based platforms which replaced Backpage; and many respondents reported that the closure of Backpage and passage of FOSTA-SESTA drove them toward street based work, agency or pimp based work, or other means of income with fewer built in safety measures (Blunt & Wolf, 2020) [3].

Independent reports refer to the St. James Infirmary, a peer-led sex worker clinic that informed us that the number of street-based workers went up by approximately 300% after the law was passed (Decriminalize Sex Work, 2023) [11]. Investigators had ceased referring trafficking cases known through online advertisements to shelters, and the alerts (advertisements) themselves had become difficult to find, as they were the main method used to identify victims and perpetrators; this made it harder to locate victims and to find missing youth, and created significant challenges for agencies to dismiss or abandon pending investigations (Decriminalize Sex Work, 2023) [11]. When the effect was described by one state investigator, it was likened to light being switched on in a room with a large number of cockroaches, with the targets fleeing in the room and investigators unable to find them out of the room (Decriminalize Sex Work, 2023) [11]. Table 6 details the key impacts of the closure of Backpage and the passage of FOSTA-SESTA.

Table 6: Documented Effects of the Backpage Closure and FOSTA-SESTA

Domain affected	Documented effect	Source
Financial stability of sex workers	More than 70% of surveyed workers reported negative financial impact	Blunt & Wolf (2020)
Access to advertising	45% of surveyed workers could not afford fee-based replacement advertising platforms	Blunt & Wolf (2020)
Street-based work exposure	Estimated tripling of street-based sex worker population post-FOSTA-SESTA	Decriminalize Sex Work (2023), citing St. James Infirmary
Investigative referral pipeline	Agencies reported an effective halt to shelter referrals sourced from online advertisements	Decriminalize Sex Work (2023)
Missing-persons and undercover investigations	Increased difficulty locating missing youth; abandonment of pending investigations	Decriminalize Sex Work (2023)

There is no clear evidence base that takedowns of platforms are generally good or bad. One specific methodological point relevant to this review is that the continued existence of the advert data that can be detected by the various types of detection systems described in Section 5.2 can be undermined by aggressive platform-level suppression, which paradoxically can decrease the number of advertisements, while also decreasing the amount of platform-level visibility for the advertisements. This trade-off needs to be considered explicitly in any operational protocol that includes both advertisement-based detection and policy advocacy; it is not necessarily a trade-off between those two actions.

6. Results: Indicators for Identifying Potential Victims and Trafficking Situations (RQ3)

6.1. Behavioral and Contextual Indicators

6.1.1. Indicators in Healthcare Settings

Trafficking victims often are not in the company of family, friends, or the public, and may appear to be an employee or acquaintance rather than a clearly distressed individual; therefore, the signs and symptoms for front-line identification are not just physical, but also contextual and behavioral (State of California Department of Justice, 2020) [41]. In reality, the majority of trafficking survivors report some contact with the healthcare system, and signs of trafficking are: the patient's sexual or reproductive health problems, or the number of partners the patient mentions, are unexpected; the patient is accompanied to the hospital or clinic by someone who will not let the patient speak for themselves or for privacy (Polaris Project, 2025) [35].

6.1.2. Indicators in Domestic and Labor Settings

Some signs of domestic servitude are an employer who keeps a worker's passport or other legal documents, and a live-in worker who sleeps in a garage, a closet, or a laundry room instead of a bedroom (Polaris Project, 2025) [35]. Labor setting indicators involve an absence of safety equipment or the presence of injuries reported to the employer, and maintaining a domestic worker's bank account, which can be conducted by the employer, and both are a shift in focus from the worker to the structure of control around him/her (Polaris Project, 2025) [35].

6.2 Digital and Advertisement-Based Indicators

6.2.1. Shared Management and Contact Overlap

In the realm of online advertising, there are several common signals that researchers have operationalized. Shared management is indicated through using third-person language (e.g., the advertisement appears to be made by someone other than the person being advertised) or by contact markers that are repeated in multiple ads that seem to describe different people, a pattern that was already noticed as a "big red flag" in 2012 by the United Nations Office on Drugs and Crime (Rodriguez Perez & Rivas, 2023) [39].

6.2.2. Mobility and Transience Language

Mobility of victims is signaled by transient-language keywords such as new in town or just arrived, reflecting both client demand for variety and traffickers' incentive to relocate operations frequently to avoid detection (Rodriguez Perez & Rivas, 2023) [39].

6.2.3. Online Recruitment and Grooming Patterns

The National Human Trafficking Hotline has seen over 950 potential sex trafficking victims who were recruited specifically online since 2015, making indicators of online recruitment not a fringe issue (Polaris Project, 2020) [33]. As per the research brief used for this review, indicative signs and characteristics are presented in Table 7, which is based on Polaris Project and governmental sources, with the intention of providing support for further investigation and not as evidence, alone, of a trafficking victim or trafficker.

Table 7: Indicators Supporting Recognition of Potential Trafficking Situations

Domain	Indicator	Source
General / offline	Employer or companion control's identity documents, wages, or living arrangements	State of California DOJ (2020); Polaris Project (2025)
Domestic servitude	Live-in worker sleeps in a non-bedroom space (garage, closet, laundry room)	Polaris Project (2025)
Healthcare setting	Patient accompanied by someone who answers for them or restricts their privacy	Polaris Project (2025)
Healthcare setting	Work-related injury with reported absence of required safety equipment	Polaris Project (2025)
Online advertisement	Identical phone number or contact marker recurs across ads depicting different individuals	Rodriguez Perez & Rivas (2023); UNODC (2012, as cited therein)
Online advertisement	Third-person language suggesting the ad was posted by someone other than the advertised person	Rodriguez Perez & Rivas (2023)
Online advertisement	Transient-location keywords (e.g., "new in town," "just arrived," "limited time")	Rodriguez Perez & Rivas (2023)
Online recruitment	Rapid escalation from social-media contact to romantic framing and requests for secrecy	Polaris Project (2020)
Financial / blockchain	Wallet clusters showing high-volume activity across multiple illicit-service categories	INTERPOL (2025b)

7. Results: Global and Regional Trends in Detected Trafficking (RQ4)

7.1. Global Prevalence Estimates

Since the 2017 edition, the ILO–Walk Free – IOM Global Estimates of Modern Slavery have provided the most methodologically sound global prevalence estimates, based on 68 nationally representative forced labour surveys, 75 forced marriage surveys and administrative data from the Counter-Trafficking Data Collaborative (ILO, 2022b) [26].

7.1.1. Forced Labor Composition and Sectoral Distribution

In 2022, it was estimated that 27.6 million people were in forced labor on any single day in 2021. Of those in forced labor, 86 percent were exploited in the private economy and 63 percent in other forms of forced labor, while 23 percent were exploited in forced commercial sexual exploitation, of which nearly four in five were women or girls, and 14 percent were a victim of state-imposed forced labor, and almost one in eight of all in forced labor were children, totalling approximately 3.3 million, including 44 percent women and girls (International Labour Organization, 2022a) [24].

7.1.2. Forced Marriage Estimates

As of 1 January 2021, an estimated 22 million people were living in forced marriage on any single day, representing 6.6 million more people than were estimated in 2016, or one in every 150 people in the world (International Labour Organization, 2022a) [24]. Forced marriage is, in most instances, a life sentence, or a long-term situation, as opposed to forced labor, which, in certain instances, can be time-limited, as reported.

7.1.3. Economic Scale of Forced Labor

The ILO's Profits and Poverty analysis estimated that worldwide, forced labour creates an estimated revenue of US$236 billion in illegal profits annually, which is equivalent to wages stolen from workers and has increased over the last 10 years due to both the growth in the number of people in forced labour and the increase in the amount of profit extracted from each individual (ILO, 2022) [24]. An ILO analysis also estimated in 2024 that it could boost GDP by around 611 billion US dollars in the global economy by ending forced labour, both by helping formerly exploited people to contribute to the economy directly and by helping to save on taxes and services further downstream.

7.2. Findings from the UNODC Global Report on Trafficking in Persons

The Global Report on Trafficking in Persons (Global TIP), as this report is known, is the 8th edition required under the UN Global Plan of Action on Trafficking in Persons (TIPS) of 2010 and is a report by UNODC (2024) that examined detected trafficking cases in 156 countries between 2019 and 2023, with summaries of over 1,000 court cases for adjudication.

7.2.1. Detection Trends, 2019-2023

There were 162 nationalities of victims trafficked into 128 destination countries in 2022, up 25 percent from 2019; forced labor victim detections rose 47 percent over the period; child victim detections rose 31 percent, and girls specifically rose 38 percent over the period; and at country level, 58 percent of child victims were trafficked within their country of origin, not across an international border (United Nations Office on Drugs and Crime, 2024) [45].

7.2.2. Regional Coverage and the African Chapter

For the first time in the Global Report series (which began in 2009), a dedicated chapter was dedicated to Africa, highlighting a correlation between climate-related displacement and the number of trafficked people detected in Europe with African origin in Africa since 2015, and a recognition that this edition includes the highest level of country coverage to date (United Nations Office on Drugs and Crime, 2024) [45].

7.3. Regional Trends Within the European Union

The annual numbers of trafficking cases reported to Eurostat offer one of the most detailed regional pictures available and depict the overall trend and composition of trafficking.

7.3.1. Registered Victims, 2021-2023

Figure 4 shows the number of registered trafficking in persons (TIP) victims in the European Union (EU) 2021-2023. The 2022 increase (to 10,093) was partly explained by the increase in awareness-raising activities linked to the influx of people fleeing war in Ukraine, while the proportion of EU citizens among victims dropped to 37 percent in 2022 compared to 2021 (European Commission, 2024) [14]. The number of registered victims increased in 2023 to 10,793, the highest since 2008 (European Commission, 2025b) [16].

Fig 4: Registered victims of trafficking in human beings, European Union, 2021–2023. Data: Eurostat (2024); European Commission (2025b).

7.3.2. Forms of Exploitation Over Time

The trend in reported forms of exploitation over the past year (2022 to 2023) is shown in Figure 5. In 2023, sexual exploitation accounted for 44 percent of registered forms of exploitation, and labour exploitation for 35 percent, while the remaining cases, such as forced criminality, forced begging and organ removal, made up just over one-fifth of registered exploitation types, reflecting a continued diversification of those types (European Commission, 2025b) [16]. For both years, women and girls accounted for around 63 per cent of registered victims. (European Commission, 2024, 2025b) [14, 16].

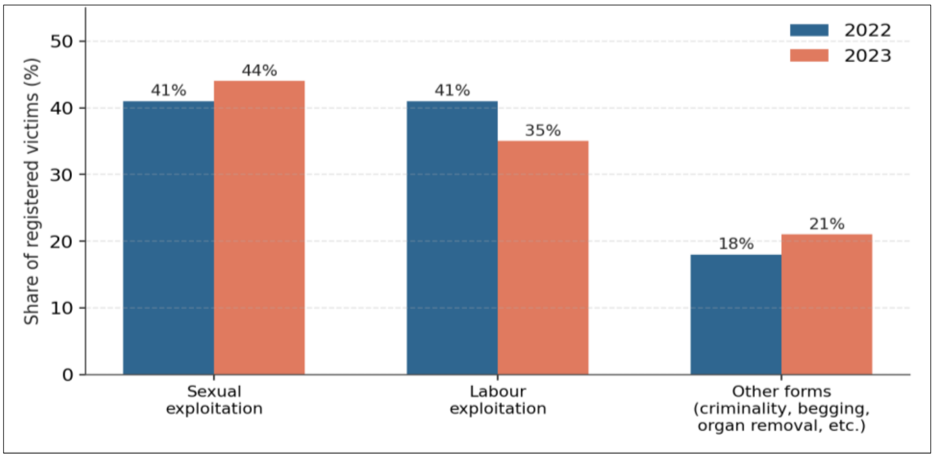

Fig 5: Reported forms of exploitation among EU-registered trafficking victims, 2022 vs. 2023. Data: European Commission, Directorate-General for Migration and Home Affairs (2024, 2025b).

7.3.3. Victim Citizenship and Cross-Border Patterns

As for the victims' citizenship, the 5th EU progress report on combating TIP reported that the number of registered victims rose by 20.5 percent compared to the 2019-2020 reporting period, with 54 percent of the victims being third-country nationals, with the top five non-EU citizenships being: Nigerian, Ukrainian, Moroccan, Colombian, and Chinese (European Migration Network Belgium, 2025) [18]. According to Figure 3, which is taken from the source data provided by the user for this review, the underlying data behind the victimisation rates by citizenship and sex for the EU27 and the ten main non-EU countries of origin is presented for 2022.

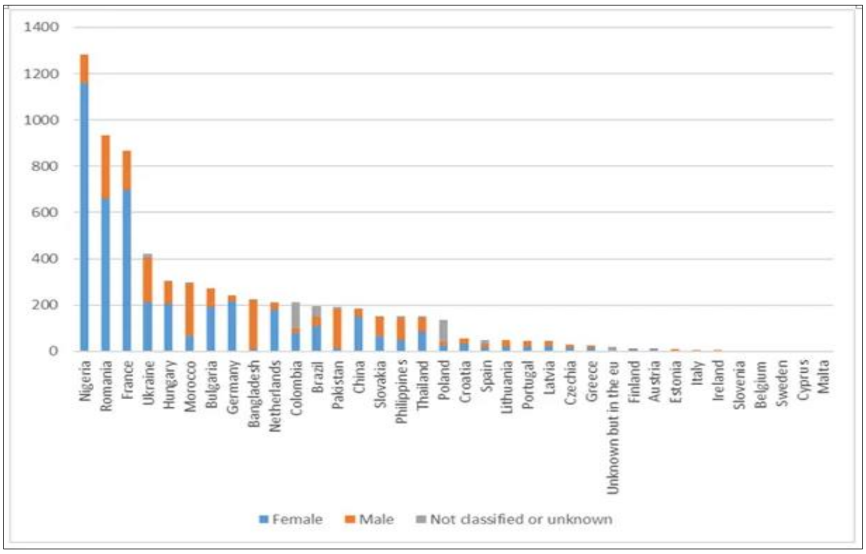

Fig 3: Number of victims per citizenship of EU27 countries and the top ten non-EU citizenships, by sex, 2022. Source: Eurostat, as presented by the European Commission, Directorate-General for Migration and Home Affairs (2024).

Table 8: Summary of Global and Regional Trafficking Statistics Cited in This Review

Indicator	Figure
Global people in forced labor (point-in-time)	27.6 million
Global people in forced marriage (point-in-time)	22 million
Children among global forced-labor population	~3.3 million (~1 in 8)
Estimated annual global profit from forced labor	US$236 billion
Estimated global GDP gain from ending forced labor	US$611 billion
Global detected victims, change 2019→2022	+25%
Global detected forced-labor victims, change 2019→2022	+47%
Global detected child victims, change 2019→2022	+31% (girls +38%)
EU registered victims, 2022	10,093 (+41% vs. 2021)
EU registered victims, 2023	10,793 (+6.9% vs. 2022)
EU registered victims, 2024	9,678
Cryptocurrency payments to suspected trafficking services, change 2024→2025	+85%
Estimated people drawn into Southeast Asian scam-compound operations	≥300,000 across ~66 countries

8. Results: Ethical, Privacy, and Bias Risks (RQ5)

8.1. Algorithmic Bias and False Identification

The same computational capacity that enables detection also enables misidentification, and the reviewed literature documents concrete, non-hypothetical instances of this risk.

8.1.1. Demographic Bias in Facial Recognition

Independent research on the facial-recognition algorithms that Clearview AI has supplied some law enforcement agencies has revealed a great deal of racial bias, with dark skin tones more likely to trigger false matches, and transgender people and women are more likely to be misidentified as men in facial-recognition systems (IEEE Computer Society VIT, 2023) [23].

8.1.2. Bias Propagation in Surveillance Analytics

On a broader level, a 2024 IBM study found that biased outcomes from AI models that analyze surveillance data can exacerbate existing privacy concerns because they are often hard to identify, and even harder to reverse, once they have occurred, which is why there are documented cases of wrongful arrests resulting from AI-assisted decision-making in law enforcement (IBM, 2024) [22]. Bias in training data and privacy concerns arising from data collection and reporting practices were found to be persistent, recurrent, and unsolved issues in the published tool ecosystem in a systematic ethicsreview of the field of AI research for trafficking, published by the Proceedings of the ACM on Human-Computer Interaction (Association for Computing Machinery, 2022) [2].

8.2. Privacy, Legal Authority, and Oversight

8.2.1. Data Governance Principles

Network analytics and dark-web surveillance is always about collecting and connecting personal information (phone numbers, images, financial identifiers, behavior patterns), and has always been argued by practitioners and ethicists to be best carried out under a defined scope of legal authority, under specific access controls, and with judicial oversight, if applicable, rather than under blanket data collection (IBM, 2024; NICE Actimize, 2025) [22, 32]. The ethics literature suggests that data-privacy protection and data bias must be addressed together in AI governance in this domain, as doing so will merely address one-half of the harm (NICE Actimize, 2025) [32].

8.2.2. Human Review Requirements

Industry analysts also specializing in cyber-forensic applications of AI have warned that AI agents trained with incomplete or biased data sets can yield misleading results, such as false positives or false negatives, that could impact the integrity of a cyber-forensic investigation if not carefully reviewed by humans at each step of its operation (Tilbury & Flowerday, 2024) [43].

8.3. Ethical Tensions Specific to Human-Rights-Centered Anti-Trafficking Research

8.3.1. Participatory Design and Survivor Involvement

Four specific suggestions for future research on how AI can help mitigate the tensions discussed above emerged from a dedicated human-rights-perspective review of AI applications to address human trafficking: (1) increased use of participatory design with trafficking survivors and affected communities; (2) increased engagement with uses of AI beyond sex trafficking, such as in labour exploitation; (3) development of shared best practices for harm prevention within research teams; and (4) transparent disclosure of AI ethics within published research, allowing downstream users to consider a tool's identified limitations before deployment (Association for Computing Machinery, 2022) [2].

8.3.2. Scope Gaps in Existing Research

The same review found that, to date, most of the published literature on how to support investigations of sex trafficking has been focused on supporting sex trafficking investigations specifically, with the other populations of trafficking survivors relatively under-served by the existing tool ecosystem (Association for Computing Machinery, 2022) [2]. The United Nations has issued a separate warning about the dangers of generative systems being used for human trafficking, such as creating convincing recruitment material, alongside the risks of generative systems to vulnerable populations in general for misinformation and information-integrity issues (United Nations, n.d.). Table 9 is a summary of the key risks found in the ethics literature that has been reviewed and the proposed safeguards to control the risks.

Table 9: Documented Risks and Corresponding Safeguards in AI-Enabled Trafficking Surveillance

Risk category	Documented manifestation	Proposed safeguard
Demographic bias in identification	Higher false-positive rates for darker-skinned individuals and gender-identification errors in facial recognition	Bias auditing prior to deployment; disaggregated accuracy reporting (IEEE Computer Society VIT, 2023)
Dataset bias and label unreliability	Trafficking advertisement labels shown to vary between human annotators	Multi-annotator agreement checks; treating classifier output as a lead, not a verdict (Cogent Social Sciences, 2024)
Privacy intrusion from data linkage	Aggregation of phone, image, and financial identifiers across platforms without consent	Restricted access, defined legal authority, judicial oversight where required (IBM, 2024)
Wrongful arrest/misidentification	Documented wrongful arrests linked to AI-assisted law-enforcement decisions	Mandatory human review before any enforcement action is taken (Tilbury & Flowerday, 2024)
Narrow research focus	Majority of tools built for sex-trafficking investigation, under-serving labor trafficking, and diverse survivor populations	Participatory design; broadened research scope (Association for Computing Machinery, 2022)
Misuse of surveillance infrastructure	Tools built for anti-trafficking purposes repurposed for unrelated monitoring	Transparent ethics disclosures; purpose-limitation policies (Association for Computing Machinery, 2022)

9. Discussion: Toward Integrated Analytic and Surveillance Protocols

9.1. Combining Structural, Textual, and Financial Evidence

The literature reviewed collectively argues for an integrated approach to network analytics and dark web surveillance, as opposed to a disjointed one. Financing is often more persistent and less likely to be altered than communications metadata: blockchain forensics provides a financial layer that is typically more persistent and difficult to fake than the communications metadata that is collected; and natural-language-processing and image-analysis systems provide the underlying ad-level data that fuels a documented network: social network analysis can help determine which actors and relations are most important in documenting a network, while natural-language-processing and image-analysis systems can be used to provide the underlying ad-level data that fuels a network at scale (Elliptic, 2025) [13]. The results of the disruption-strategy comparison (summarized in Table 2) indicate that the choice of which type of actors to target (social-capital brokers, human-capital specialists, or financial coordinators) should not be left to a default targeting approach, but should instead be based on the specific structural role of the network in question. The platform-policy findings in Section 5.5 bring additional complexity: aggressive takedown efforts may have a negative impact on the digital visibility that forms the basis of these analytic tools, so it is important to consider the potential loss of digital visibility while assessing the value of detection.

9.2. A Proposed Operational Protocol

The reliability concerns described in Section 5.2.6, and the bias and privacy concerns described in Section 8 also argue for the need to use all such outputs as investigative leads to be followed up and corroborated – as this study has framed it, indicators should be used to inform further investigation but should not, in isolation, be used to establish a victim or an offender. Part of this operational protocol, built on the resources reviewed, would be legally authorized data collection, corroboration across text, image, network and financial evidence, human review at every escalation level, and a clear protocol for referring identified potential victims to healthcare, shelter, legal and rehabilitation services, while not treating the act of identification as a "done deal," but rather as a tool for referring patients to services (Zimmerman & Kiss, 2017) [55].

9.3. Answering the Research Questions

RQ1 sought to understand how network analysis can detect structurally important actors and break the trafficking networks. The results in Section 4 confirm that measures of centrality (degree, betweenness, eigenvector, and closeness) and community-detection methods accurately identify a few structurally important actors, and that the removal of these actors, based on social capital, human capital, or financial-coordination role, is more effective than random removal, while the networks have some ability to adjust and reconstitute after these actors are removed.

RQ2 was concerned with the ability of dark-web surveillance, NLP (Natural Language Processing), image analysis, and blockchain forensics to identify advertisements and financial transactions associated with trafficking. The progression in technical sophistication from keyword-based classifiers towards multimodal, stylometric, and financially related detection methods is clearly visible in Section 5, in particular, blockchain forensics provides a hard to fabricate evidentiary channel; however, the potential of undermining the very digital visibility that these methods rely on was not previously made explicit in either computer-science or policy literature reviewed, as this is the case with platform-level suppression of advertising in Section 5.5.

RQ3 was: What are the indicators to support identification of potential victims and trafficking situations? Section 6 pulls together behavioural, contextual, and digital-advertisement data from both governmental and practitioners, and none of these data points alone is enough to substantiate a victim or an offender without further investigation.

RQ4 sought to determine what information global and regional data show about recent trafficking trends. Even though trafficking is an illegally, unethically, and immorally traded victimization experience, and that often only a small number of such individuals are being trafficked, there is also a steady rise in the number of trafficked victims in both global (ILO, UNODC) [24, 45] and regional (Eurostat, European Commission) [19, 16] data sources, with a documented diversification of types of trafficking and a continued dominance of female and female children among registered victims.

RQ5 aimed to identify ethical, privacy, and bias challenges that are associated with these techniques, and to identify protections to address those challenges. Section 8 summarizes the literature reviewed, highlighting the presence of demographic bias, unreliability of dataset labels, intrusion on privacy, and risk of wrongful identification, as well as a common set of proposed safeguards: bias auditing, restricted access and authorized access, mandatory human review, and participatory research design involving survivors.

10. Limitations and Recommendations

10.1. Methodological Limitations

This study has several limitations that mirror those of the underlying literature it synthesizes. The bulk of the network-analytic case-study evidence comes from a few jurisdictions, primarily the United Kingdom and the United States, and the different results in the case studies between the United Kingdom and Latino-network (Section 4.4) indicate that structural findings might not transfer cleanly across regions and across different types of exploitation and cultural contexts. The literature on machine learning detection presented in Section 5.2 is an ongoing methodological discussion on label reliability, and this study has not settled the issue, but rather furthered the discussion. The platform-policy evidence presented in Section 5.5 relies, in large part, on a single, sex-worker-led survey, with a small (but independently corroborated by journalistic and advocacy reporting) combined sample size of 136 respondents; it should not be interpreted as a representative national estimate of the impacts of FOSTA-SESTA. This study is a qualitative document-analysis design, rather than a systematic review protocol, and so the selection of sources was not systematic and exhaustive, nor were quantitative synthesized meta-analytic estimates (such as pooled detection accuracy across studies) included.

10.2. Scope Limitations

The Global and regional prevalence statistics presented in section 7 only refer to cases detected and registered, and the detection capacity varies by country and over time (United Nations Office on Drugs and Crime, 2024; Eurostat, 2024) [45, 19]. This study did not gather any original network data, survey data, or interview data, and all reported effect sizes and accuracy percentages and outcomes of cases are inherited from the cited primary sources and were not re-checked for the accuracy of the data beyond comparing with a second source when available.

10.3. Recommendations for Future Research

Building on the gaps identified above, future research would benefit from four developments. First, if the centrality-based disruption studies reviewed in Section 4 were replicated in a broader range of jurisdictions and different types of exploitation, then some of the structural findings would be better able to determine whether they are generalizable or specific to certain jurisdictions. Second, the datasets used to train the classifiers reviewed in Section 5.2 for the ads we examined would be relabeled separately and by multiple annotators, directly addressing the label-reliability critique from Section 5.2.6. Third, the impact of platform-policy interventions would be better addressed by a prospective, controlled evaluation – Ideally comparing jurisdictions that did, with those that did not, adopt liability rules like FOSTA-SESTA. Fourth, participatory research with trafficking survivors that is explicitly engaged in designing and evaluating tools is not yet commonly found in the literature, addressing the scope gaps identified in Section 8.3.2, as recommended by the Association for Computing Machinery (2022) [2].

11. Conclusion

Network analytics and lawful dark-web surveillance have developed, in the last decade or so, from mostly conceptual proposals into a set of operational tools that have been used to achieve specific investigative results: centrality-guided targeting has been used to inform real prosecutions, advertisement-classification systems have been used to detect victims that had previously gone undetected, and blockchain forensics has been used to trace tens of millions of dollars of financial flows known to be connected to trafficking, to their cash-out point. Concurrently, the same literature is remarkably frank about the constraints and dangers of these tools, including the known demographic bias, label unreliability, privacy problems, and even the potential that much-needed policy intervention in the platforms can erode the digital visibility that these methods rely on. What the evidence reviewed suggests is that these technologies cannot replace human trafficking efforts to address it, but rather that they can provide a way to generate structured intelligence from the myriad and uncertain digital traces, if they are designed from the beginning with legal authorization, human oversight, and pathways to refer to survivors.

References

1. Antonopoulos GA, Baratto G, Di Nicola A, Diba P, Martini E, Papanicolaou G, *et al*. Technology in human smuggling and trafficking. Springer; 2020.
2. Association for Computing Machinery. Ethical tensions in applications of AI for addressing human trafficking: a human rights perspective. Proc ACM Hum Comput

Interact. 2022.
3. Blunt D, Wolf A. Erased: The impact of FOSTA-SESTA and the removal of Backpage on sex workers. Anti-Trafficking Rev. 2020;(14):117-121. doi:10.14197/atr.201220148.
4. Chainalysis. Combating trafficking with blockchain analysis. Chainalysis Blog. 2022 Apr 12. Available from: https://www.chainalysis.com/blog/combatting-trafficking-with-blockchain-analysis/
5. INTERPOL. INTERPOL releases new information on globalization of scam centres. 2025 Jun 30. Available from: https://www.interpol.int/en/News-and-Events/News/2025/INTERPOL-releases-new-information-on-globalization-of-scam-centres
6. INTERPOL. Growing threat of transnational scam centres highlighted at INTERPOL General Assembly. 2025 Nov 26. Available from: https://www.interpol.int/en/News-and-Events/News/2025/Growing-threat-of-transnational-scam-centres-highlighted-at-INTERPOL-General-Assembly
7. Financial Crimes Enforcement Network. FinCEN issues final rule imposing special measure five against Huione Group. U.S. Department of the Treasury. 2025 Oct 15. Available from: https://www.fincen.gov/news/news-releases/fincen-issues-final-rule-imposing-special-measure-five-against-huione-group
8. Cockbain E, Brayley H, Laycock G. Exploring internal child sex trafficking networks using social network analysis. Polic J Policy Pract. 2011;5(2):144-157. doi:10.1093/police/par027.
9. Cogent Social Sciences. Why we cannot identify human trafficking from online advertisements. Cogent Soc Sci. 2024. doi:10.1080/23322705.2024.2435198.
10. Duijn PAC, Kashirin V, Sloot PMA. The relative ineffectiveness of criminal network disruption. Sci Rep. 2014;4:4238. doi:10.1038/srep04238.
11. Decriminalize Sex Work. What is SESTA/FOSTA? 2023. Available from: https://decriminalizesex.work/advocacy/sesta-fosta/what-is-sesta-fosta/
12. DTR Society. AI-enhanced detection of human trafficking. 2025. Available from: https://www.dtrsociety.org/wp-content/uploads/library/nice2025/articles/ainow2025_20044.pdf
13. Elliptic. How to squeeze investigative evidence from any cryptocurrency address. Elliptic Blog. 2025 Sep 10. Available from: https://www.elliptic.co/blog/how-to-squeeze-investigative-evidence-from-any-cryptocurrency-address
14. European Commission, Directorate-General for Migration and Home Affairs. Newly released data show an increase of trafficking in human beings. 2024 Feb 28. Available from: https://home-affairs.ec.europa.eu/news/newly-released-data-show-increase-trafficking-human-beings-2024-02-28_en
15. European Commission, Directorate-General for Migration and Home Affairs. New progress report on combatting trafficking in human beings. 2025 Jan 20. Available from: https://home-affairs.ec.europa.eu/news/new-progress-report-combatting-trafficking-human-beings-2025-01-20_en
16. European Commission, Directorate-General for Migration and Home Affairs. New data indicates an increase of victims of trafficking in human beings in the EU. 2025 Apr 7. Available from: https://home-affairs.ec.europa.eu/news/new-data-indicates-increase-victims-trafficking-human-beings-eu-2025-04-07_en
17. European Commission. Statistics and trends in trafficking in human beings in the European Union, 2021-2022, accompanying the fifth progress report (SWD(2025) 4 final). EUR-Lex. 2025. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52025SC0004
18. European Migration Network Belgium. The share of non-EU nationals among trafficking victims in the EU is increasing, according to the EU Commission. 2025 Jan 20. Available from: https://emnbelgium.be/news/share-non-eu-nationals-among-trafficking-victims-eu-increasing-according-eu-commission
19. Eurostat. Trafficking in human beings' statistics. Statistics Explained. 2024. Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Trafficking_in_human_beings_statistics
20. Summers L, Shallenberger AN, Cruz J, Fulton LV. A multi-input machine learning approach to classifying sex trafficking from online escort advertisements. Mach Learn Knowl Extr. 2023;5(2):461-476. doi:10.3390/make5020028.
21. Ficara A, Curreri F, Fiumara G, De Meo P. Human and social capital strategies for Mafia network disruption. arXiv. 2022. Available from: https://arxiv.org/pdf/2209.02012
22. IBM. Exploring privacy issues in the age of AI. IBM Think. 2024. Available from: https://www.ibm.com/think/insights/ai-privacy
23. IEEE Computer Society VIT. The ethics of AI: addressing bias and privacy concerns. Medium. 2023 Jun 21. Available from: https://medium.com/@IEEE_Computer_Society_VIT/the-ethics-of-ai-addressing-bias-and-privacy-concerns-573df8ca1490
24. International Labour Organization, Walk Free, International Organization for Migration. Global estimates of modern slavery: forced labour and forced marriage. Geneva: International Labour Organization; 2022. Available from: https://publications.iom.int/books/global-estimates-modern-slavery-forced-labour-and-forced-marriage
25. International Labour Organization. 50 million people worldwide in modern slavery. ILO News. 2022 Sep 12. Available from: https://www.ilo.org/resource/news/50-million-people-worldwide-modern-slavery-0
26. International Labour Organization. Global estimates of modern slavery: forced labour and forced marriage—methodology. 2022. Available from: https://www.ilo.org/publications/global-estimates-modern-slavery-forced-labour-and-forced-marriage-1
27. International Labour Organization. Global efforts to end forced labour could increase global GDP by US$611 billion. 2024 Sep 18. Available from: https://www.ilo.org/resource/news/global-efforts-end-forced-labour-could-increase-global-gdp-us611-billion
28. Klabbers RE, *et al*. Human trafficking risk factors, health impacts, and opportunities for intervention in Uganda: a qualitative analysis. Glob Health Res Policy. 2023;8:52.

29. Medipelly S, Abosata N. Detection of dark web threats using machine learning and image processing. arXiv. 2024. Available from: https://arxiv.org/pdf/2407.00704
30. MIT News. Artificial intelligence shines light on the dark web. 2019 May 13. Available from: https://news.mit.edu/2019/lincoln-laboratory-artificial-intelligence-helping-investigators-fight-dark-web-crime-0513
31. National Center on Safe Supportive Learning Environments. Indicators. U.S. Department of Education. Available from: https://safesupportivelearning.ed.gov/human-trafficking-americas-schools/indicators
32. NICE Actimize. The ethics of AI in monitoring and surveillance. 2025 Jun 20. Available from: https://www.niceactimize.com/blog/fmc-the-ethics-of-ai-in-monitoring-and-surveillance
33. Polaris Project. Human trafficking and online recruitment. 2020 Jan 6. Available from: https://polarisproject.org/human-trafficking-and-online-recruitment/
34. Polaris Project. Recognizing labor trafficking. 2024 Feb 29. Available from: https://polarisproject.org/labor-trafficking/
35. Polaris Project. Recognizing human trafficking. 2025 Dec 8. Available from: https://polarisproject.org/recognizing-human-trafficking/
36. Polaris Project. Recognizing sex trafficking. Available from: https://polarisproject.org/sex-trafficking/
37. Portnoff RS, Huang DY, Doerfler P, Afroz S, McCoy D. Backpage and Bitcoin: uncovering human traffickers. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2017. p.1595-1604. doi:10.1145/3097983.3098082.
38. ResearchGate. Human trafficking and the darknet: technology, innovation, and evolving criminal justice strategies. 2020. Available from: https://www.researchgate.net/publication/342254004
39. Rodriguez Perez A, Rivas P. Combatting human trafficking in the cyberspace: a natural language processing-based methodology to analyze the language in online advertisements. arXiv. 2023. Available from: https://arxiv.org/pdf/2311.13118
40. ShadowDragon. Spotting the signs: how social media and the dark web can reveal human trafficking activity. 2024 Jul 15. Available from: https://shadowdragon.io/blog/spotting-the-signs-how-social-media-and-the-dark-web-can-reveal-human-trafficking-activity/
41. State of California, Department of Justice, Office of the Attorney General. Identifying human trafficking. 2020 Jan 15. Available from: https://oag.ca.gov/human-trafficking/identify
42. Tasleem N, Ansari MN, Raghav RS. Data gravity vs. model agility: the new tension shaping the future of automation and AI: a systematic review. J Inf Syst Eng Manag. 2025;10(49s):1041-1056.
43. Tilbury J, Flowerday S. Humans and automation: augmenting security operation centers. J Cybersecur Priv. 2024;4(3):388-409. doi:10.3390/jcp4030020.
44. Tong E, Zadeh A, Jones C, Morency LP. Combating human trafficking with multimodal deep models. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics; 2017. p.1547-1556. doi:10.18653/v1/P17-1142.
45. United Nations Office on Drugs and Crime. UNODC global human trafficking report: detected victims up 25 per cent as more children are exploited and forced labour cases spike. 2024 Dec 11. Available from: https://www.unodc.org/unodc/en/press/releases/2024/December/unodc-global-human-trafficking-report_-detected-victims-up-25-per-cent-as-more-children-are-exploited-and-forced-labour-cases-spike.html
46. United Nations. Modern slavery is on the rise. 2022 Sep 12. Available from: https://www.un.org/en/node/111996
47. United Nations. Safeguarding human rights and information integrity in the age of generative AI. UN Chronicle. Available from: https://www.un.org/en/UN-chronicle/safeguarding-human-rights-and-information-integrity-age-generative-AI
48. Mapping trafficking networks: a data-driven approach to disrupt human trafficking post Russia-Ukraine conflict. arXiv. 2025. Available from: https://arxiv.org/pdf/2504.17050
49. Masud N, Al Bassam A, Sobhan A, Sourav MTI. Data-driven solutions for human trafficking: detection and impact analysis using machine learning. In: 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE); 2025. IEEE. Available from: https://ieeexplore.ieee.org/document/11014018/
50. Vajiac C, Chau DH, Olligschlaeger A, Mackenzie R, Nair P, Lee MC, *et al*. TrafficVis: visualizing organized activity and spatio-temporal patterns for detecting and labeling human trafficking. IEEE Trans Vis Comput Graph. 2022.
51. Vajiac C, Lee MC, Kulshrestha A, Levy S, Park N, Olligschlaeger A, *et al*. DeltaShield: information theory for human-trafficking detection. ACM Trans Knowl Discov Data. 2023;17(2). doi:10.1145/3563040.
52. Vanderbilt Institutional Repository. Approach to the global human trafficking crisis: analyzing applications of social network analysis. 2022 Mar 27. Available from: https://ir.vanderbilt.edu/items/5b5eb3ec-02e6-4ff0-8f92-fb7c568f8c8e
53. Office of the United Nations High Commissioner for Human Rights. Online scam operations and trafficking into forced criminality in Southeast Asia: recommendations for a human rights response. 2023 Aug 29. Available from: https://bangkok.ohchr.org/reports/online-scam-operations-and-trafficking-forced-criminality-southeast-asia-recommendations
54. Castellanos M. U.S. seizes $15 billion in bitcoin, sanctions Cambodia's Prince Group in global crypto scam crackdown. Forbes. 2025 Oct 14. Available from: https://www.forbes.com/sites/martinacastellanos/2025/10/14/us-seizes-15-billion-in-bitcoin-sanctions-cambodias-prince-group-in-global-crypto-scam-crackdown/

55. Zimmerman C, Kiss L. Human trafficking and exploitation: a global health concern. PLoS Med. 2017;14(11):e1002437. doi:10.1371/journal.pmed.1002437.
56. Zimmerman C, Hossain M, Watts C. Human trafficking and health: a conceptual model to inform policy, intervention and research. Soc Sci Med. 2011;73(2):327-335. doi:10.1016/j.socscimed.2011.05.028.

How to Cite This Article
Abah O. The Role of Network Analytics and Dark-Web Surveillance in Identifying Potential Victims, Tracking Human-Trafficking Trends, and Disrupting Organized Trafficking Networks. J Front Multidiscip Res. 2026;7(2):137-154. doi:10.54660/.JFMR.2026.7.2.137-154