A Machine Learning and Theoretical Approach to Understanding Technology-Facilitated Disinformation and Misinformation in Human Trafficking

Student number: 23220031

Master Of Sicence in AI &Ethics

Supervisors: Mahsa Abazari, Adrian Hillman, Ioannis Kypraios

A Dissertation

Presented to the Northeastern University London

A Machine Learning and Theoretical Approach to Understanding Technology-Facilitated Disinformation and Misinformation in Human Trafficking

Supervisors: Mahsa Abazari, Adrian Hillman, Ioannis Kypraios

In this dissertation, we aim to understand how the development of the World Wide Web and social networks has allowed unprecedented information dissemination across the general population and facilitated the proliferation of false and misleading information on a larger scale. Our intention is to analyze the harm caused by disinformation and misinformation, respectively. Misinformation refers to false information, such as fake news, that spreads online and through other channels. This often occurs due to a lack of authentication processes in online publishing or the ease and speed of sharing information on social media, where users may post without verifying facts or while under emotional influence.

Disinformation, on the other hand, occurs when false information is spread intentionally to harm or deceive. With the recent proliferation of artificial intelligence technology, especially generative AI and large language models such as ChatGPT, new avenues for the dissemination of misinformation and disinformation have emerged, including through videos, texts, images, and audio. This technology-facilitated misinformation and disinformation expose the public to threats from organized crime organizations, which can use misinformation to elude law enforcement or target vulnerable populations by spreading disinformation for specific crimes, such as human trafficking, particularly through the recruitment of victims into forced labor by spreading false job advertisements online.

The dissertation employs a systematic literature review to identify human trafficking indicators and the strategy of deliberately spreading disinformation through false job advertisements to recruit potential victims. This is followed by the creation of human trafficking-specific keywords/phrases related to such strategies, and then using the vocabularies to generate datasets for machine learning model training. The findings suggest there is potential for operationalizing the use of human trafficking-related indicators in training these models. However, the complex social, economic, and cultural factors surrounding human trafficking issues require that technological solutions pay more attention to the design of research methodologies.

Forced labor victimization has risen over the years, refer to Figure 1. We call for urgent interventions from various stakeholders, including the government, policymakers, social media platforms, and the technological sector. Additionally, it is crucial to devise effective countermeasures utilizing machine learning or deep learning algorithms to detect and discern misinformation and disinformation.

Figure 1.0

Available from: https://ec.europa.eu/eurostat/statistics-explained/images/0/04/Trafficking_in_human_beings_statistics_Highlight_24-01-2024.png

Commented [JH1]: Table 1

Contents 3

Abstract

Chapter 1 Introduction 4

1.1 Motivation

1.2 Dissertation Structure

1.3 Research Questions & Objectives

1.4 Project Plan

Chapter 2 Literature Review 7

2.1 Misinformation and Disinformation

2.2 Drivers and Countermeasures of Technology-Facilitated Misinformation (TF- MI)

2.2.1 The Drivers of Technology-Facilitated misinformation (TF-MI)

2.2.2 Current Countermeasures of TF-MI

2.3 Drivers and Countermeasures of Technology-Facilitated Disinformation (TF-DI)

2.3.1 The Drivers of Technology-Facilitated Disinformation (TF-DI)

2.3.2 Current Countermeasures to TF-DI

2.4 Misinformation Risks Posed by Generative Models

2.5 Disinformation Risks Posed by Generative Models

2.6 Case Studies:

2.6.1 Misinformation & Disinformation: Human Trafficking

2.6.2 Current Countermeasures to Technology Facilitated Trafficking of Human Being (TF-THB)

2.6.3 Current Countermeasures in (TF-THB): Detect Fraudulent Job Advertisement

2.6.4 Methodological challenges around HT research

Chapter 3 Methodology

3.1 Research Methodology

3.2 Implementation methodology

3.3 Research Ethics

Chapter 4 Implementation

4.1 Sampling and Data

4.1.1 Literature Review on Human Trafficking Detection Indicators

4.1.2 Selection of Key Words & Phrases

4.1.3 Human Trafficking related Job Advertisements Dataset Vocabulary (By categories)

4.1.4 Legal Job Advertisements Dataset Vocabulary (By categories)

4.2 Data Generation

4.3 Model Implementation

4.3.1 Data Preparation

4.3.2 Baseline Model Training: Logistic Regression

4.3.3 Advanced Model Training: Support Vector Machine

Chapter 5 Findings

5.1 Model Findings and Comparisons

5.1.1 Model performance on Unseen datasets

5.2 Dataset Characteristics: Class 1 and Class 0

5.3 Application of models in Real World Data

Chapter 6 Discussion

6.1 Methodological challenge

6.2 Operationalize Huamn trafficking Indicators

6.3 Sociological Factors

6.4 Technology Facilitated Disinformation Strategy and in relations to Human Trafficking (recruitment)

Conclusion

Chapter 1 Introduction

1.1 Motivation

Misinformation and disinformation have been widespread global concerns, causing economic harm and hindering political and social justice. The rise of World Wide Web technology and social media platforms allows for quicker and easier dissemination of false information. On top of that, recent technological advancements in generative AI models such as ChatGPT, Sora, and Stable Diffusion enable new methods of generating content, aggregating Artificial Intelligence Generated Content (AIGC) [7]. We have seen a rise in technology-facilitated misinformation and disinformation, driven by the same motivations behind traditional misinformation, where false information is spread with the intention to harm. With the rise of generative AI models and the widespread use of social media platforms by billions around the world, these tools are further utilized to aggregate misinformation and disinformation. This presents new challenges from technological, sociological, and criminological perspectives in detecting false content and understanding the motivations behind technology-facilitated misinformation and disinformation.

Therefore, in this dissertation, the motivation is to analyze how technology has facilitated the further generation of misinformation and disinformation from a technological point of view and to find a technological solution to detect both misinformation and disinformation. This then guides us into understanding how organized crime organizations utilize misinformation and disinformation to commit crimes, specifically human trafficking. By understanding their technological and criminological motivations, we aim to devise a feasible technological solution to combat misinformation and disinformation threats in the context of deterring human trafficking.

1.2 Research Questions & Objectives

RQ1: How do organized crime syndicates utilize misinformation (MI) and disinformation (DI) effects and tactics for human trafficking (HT)?

RQ2: How can the state-of-the-art detection and intervention framework for MI and DI be combined to apply them for the detection and prevention of online HT activities?

Objectives:

- Conduct a literature review on the topics of:
    1. Misinformation and disinformation
    2. Countermeasures

3. Human trafficking activities and implications with MI and DI
4. Countermeasures
    - Set a theoretical framework
    - Collect data
    - Implement solutions
    - Discuss findings

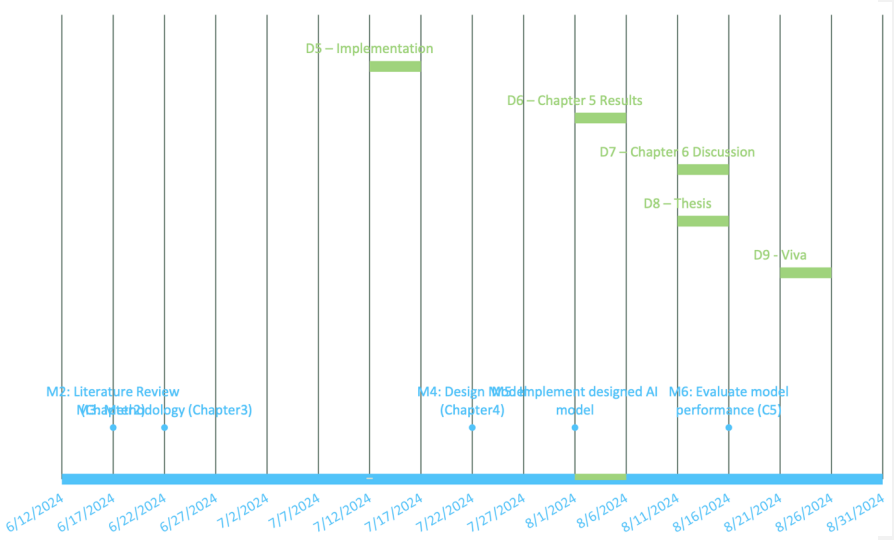
1.3 Dissertation Structure

This dissertation consists of seven chapters. Chapter 1 introduces the motivation for the project. Chapter 2 provides a literature review, divided into six sections to establish the theoretical foundation of technology-facilitated misinformation and disinformation, the technological solutions (machine learning) to combat both, and case studies on how human trafficking detection and prevention are threatened by technology-facilitated misinformation and disinformation, as well as the methodological challenges around human trafficking research. Chapter 3 outlines the research and application methodology. Chapter 4 delves into the implementation and data sampling process for machine learning training to combat technology-facilitated disinformation in human trafficking. Chapter 5 discusses the findings and the deployment of the trained model on real-world data. Chapter 6 explores potential future improvements and additional observations. Finally, Chapter 7 provides concluding remarks.

1.4 Project Plan

Below, in Figure 2, a Gantt chart visualizes the timelines for milestones and deliverables in achieving this dissertation. The dissertation aims to answer RQ1 and RQ2 in two stages. First, ensuring adherence to ethical guidelines and reviewing the dissertation process. Second, conducting a literature review and submitting a research proposal that clearly states the direction and plan for achieving the objectives. Third, gathering data and implementing machine learning techniques to answer RQ2. Fourth, revising implementation results and constructing a theoretical framework with practical interpretation for RQ1. Finally, submitting the final dissertation and preparing to present the VIVA. The actual VIVA has been moved to Monday, the 2nd of September, and the submission deadline has been extended to the 30th of August.

Figure 2.0


D5 – Implementation
D6 – Chapter 5 Results
D7 – Chapter 6 Discussion
D8 – Thesis
D9 - Viva
M2: Literature Review (Chapter2)
M3: Methodology (Chapter3)
M4: Design Model (Chapter4)
M5: Implement designed AI model
M6: Evaluate model performance (C5)
6/12/2024
6/17/2024
6/22/2024
6/27/2024
7/2/2024
7/7/2024
7/12/2024
7/17/2024
7/22/2024
7/27/2024
8/1/2024
8/6/2024
8/11/2024
8/16/2024
8/21/2024
8/26/2024
8/31/2024

Chapter 2 Literature Review

2.1 Misinformation and Disinformation

Misinformation is defined as false or incorrect information, sometimes interchanged with the term ‘fake news’ [1]. This dissertation exclusively refer to it as ‘misinformation.’ It is recognized as a particular challenge that harms and affects society in various ways, causing damage to economic and political stability and clouding individual judgment and public policy decisions [2]. Misinformation often occurs due to unverifiable sources that fail to authenticate the truthfulness of the information. This issue is exacerbated by the increasing number of outlets for spreading information and news, expanding from a few to virtually everyone due to the advancement of social networks [3], namely technology-facilitated misinformation (TF-MI). This leads to a social phenomenon that increases public polarization and distrust in politics and journalism [4]. However, the intent of the agent generating and sharing false information differentiates disinformation, where the agent deliberately shares inaccurate information with the intention to harm and achieve malicious goals, from misinformation, where the agent unconsciously spreads false news [5].

Disinformation can be organized by both state and non-state actors, including individuals or organized groups. It is created, spread, and amplified both organically by exploiting cognitive biases such as attentional and confirmation biases, and through the use of astroturfing techniques (bandwagon effect) by creating the impression of widely shared beliefs around a particular issue or item. Such disinformation campaigns can target and discredit verifiable news channels or those who hold opposing views, with disinformation agents spreading such campaigns using threats, intimidation, and disruptive tactics [5]. Disinformation has similarly worsened due to the advancement of the World Wide Web, namely technology-facilitated disinformation (TF-DI).

2.2 Drivers and Countermeasures of Technology-Facilitated Misinformation (TF-MI)

2.2.1 The Drivers of Technology-Facilitated Misinformation (TF-MI)

This overview looks at previous works on TF-MI, categorizing what could facilitate or drive it into two sectors: human-centric and socio-technological centric.

The human-centric explanations focus on the human aspect of why people are prone to share or re-share misinformation online. ‘Fake news travels faster than true stories.’ Quoting three MIT scholars’ research findings, they found that after eliminating fake news disseminated by bots, humans are overwhelmingly the primary agents spreading fake news on Twitter, not the bots. Fake news is 70% more likely to be shared than true stories. True stories take six times longer to reach 1500 people than fake news [6]. In the work by Danni Xu et al., they particularly take a social science perspective, examining psychological, behavioral, and interpersonal factors [7]. The psychological drivers of misinformation can be accounted for by the information deficit model. This inherently points towards the deficiency of reliable, responsible, and truthful data provided to the general public, leading to public misconception and lack of truthful knowledge. However, the drivers of misinformation are multifaceted. Another key factor the authors mention is confirmation bias—when individuals seek information that confirms their beliefs and disregard information that opposes said beliefs.

Another factor to consider is the information glut—the abundance of information available can lead to difficulties in seeking truth and making decisions, resulting in individuals defaulting to the information outlet that is easiest to reach. Additionally, affective elements—people's emotions and moods when they perceive misinformation—are generally considered a factor in distinguishing misinformation. In Danni Xu et al., they brought up that political tweets involving emotional content have a 20% higher likelihood of being re-tweeted. Lastly, the continued influence effect (CIE), the residual effect of misinformation that is not corrected and leaves individuals continually consuming false information based on the false channel [7].

Moving on to the socio-technological factors, this encompasses how broader societal and technologically enabled environments can drive misinformation. In Pareek and Goncalves's work, they examined how social media platforms' engagement-driven algorithms facilitate further misinformation. The platforms are designed to prompt users to click on posts by making them attractive. Additionally, social media platforms, such as Facebook, which are profit-driven, need to obtain more clicks from users to generate revenue. When truth-labels supplied by the platform are resisted due to users' distrust of institutions, misinformation posts thrive better and receive more clicks than truthful posts [4]. Additionally, when a crisis occurs, misinformation tends to increase. For example, during COVID-19, numerous untrue contents spread online about health-related issues, leading to public distrust in health policy advice from institutions. This is because, during crises, social media platforms allow fast and prompt outlets for people to post content, and it is dire to eliminate the misinformation from the truthful content to provide informed decisions to the institutions and the public [1].

2.2.2 Current Countermeasures of TF-MI

Reviewing the current efforts in mitigating misinformation content. The prominent techniques include utilizing machine learning methods to detect 'fake news' content [3], detecting misinformation from the perspective of emotion analysis combined with user behavioral analysis [8], and from the perspective of content novelty combined with emotion analysis [9]. Besides detection methods, a novel intervention technique involves introducing peer-supplied credibility labels instead of those from social platforms to increase convincing effects on users [4].

Detection techniques focus on two factors: detecting fake news content and detecting human-centric factors combined with content detection. In Sudhakar and Kaliyamurthie, the authors use various machine learning methods and deep learning classifiers to improve the detection system's ability to distinguish between truthful information and falsehoods, using a dataset containing tweets related to COVID-19. The authors use machine learning and deep learning techniques: Support Vector Machine (SVM), Naïve Bayes (NB), Logistic Regression (LR), and Recurrent Neural Networks (RNN). They noted the highest performance of SVM and NB classifiers. The best performance yielded from the combination of the Term Frequency-Inverted Document Frequency (TF-IDF) technique with a Linear SVM classifier with 98% accuracy. Noticeably, the authors propose combining the use of metadata and text analysis to optimize detection system accuracy. Metadata offers additional contextual understanding that textual data alone does not provide. Essentially, it summarizes and gives basic information about the data [10]; in the case of fake news detection, it can perhaps contain and extract information such as publication dates and authors' names, thus providing further background and contextual information. It can also be applied to Natural Language Processing (NLP) to

analyze the text of news articles to identify patterns and markers that may indicate falsehoods [3].

Jouhar et al. (2024) utilized the ISOT fake news dataset and TF-IDF feature extraction technique. They pre-processed text data and chose models for classification purposes to classify news articles into two categories: real/fake. The models included Logistic Regression (LR), Decision Tree, Random Forest, Gradient Boosting, XGBoost, and Passive Aggressive Classifier. With LR as the baseline model, all models were evaluated on accuracy, precision, recall, F1-score, and AUC (Area Under the Curve). XGBoost gave the best performance across all metrics due to its ability to handle non-linear relationships and regularization techniques, showing minimal overfitting and good generalization. The authors suggest future research should explore parameter tuning and additional data vectorization methods such as Word2Vec and BERT [11].

The detection techniques also investigate human-centric factors. In Indu V. and Sabu M. Thampi (2024), they leverage emotion analysis and user behavior attributes to enhance misinformation detection accuracy by utilizing a Recurrent Neural Network (RNN) to explore users' emotional attributes expressed in social network texts for the detection of misinformation. Additionally, they combine six user behavioral attributes extracted from users' social media profiles using a Fuzzy inference system to analyze user attributes. The authors noted the significance of combining users' emotions and user attributes—presence of URLs in tweets, account registration time, follower count, following count, account age, and retweet count—to analyze the likelihood of misinformation. They tested on five real-world datasets and found that tweets exhibiting higher emotional intensity, especially anger and fear, tend to contain misinformation. Real information showed a more balanced emotional distribution. Similarly, they found that the six behavior attributes contribute some patterns and markers that can be combined with emotional analysis to increase detection accuracy. The results show that for each dataset, the majority of misinformation tweets were correctly classified into the higher probability categories, validating the effectiveness of the combined emotion and user behavioral approach [8].

Another approach proposed by Kumari et al. (2021) combines novelty detection and emotional analysis incorporated into fake news detection. It uses a deep learning-based multitask learning framework to perform three tasks. The authors propose that the mobility and emotional content significantly attribute to the virality of fake news. The implementation is in two phases: 1. Generating ground truth novelty and emotion labels using trained models on datasets like Quora Question Pairs and GoEmotion. 2. Using the labels to train the multitask model using shared representations among the three tasks. The results conducted on datasets such as ByteDance, FNC, Covid-Stance, and FNID demonstrated improvements over baseline models [9].

Aside from detection methods, novel intervention techniques focus on supplying truth labels on content. In Pareek and Goncalves (2024), they proposed an intervention technique to address misinformation by exploring the impact of peer-supplied credibility labels. They used a dataset that included user behavior attributes such as emotional tone, engagement level, and sharing patterns, employing machine learning models to assess effectiveness. The study shows that peer-supplied credibility labels significantly increase the accuracy of misinformation detection. They found that it mitigates 'confirmation bias' to a degree, as participants' belief in misinformation significantly decreased when peers from the same political group disputed the credibility, which corresponds to the in-group dynamics.

Noticeably, belief in misinformation was reduced more effectively when the credibility dispute came from peers with high political agreement [4].

2.3 Drivers and Countermeasures of Technology-Facilitated Disinformation (TF-DI)

2.3.1 The Drivers of Technology-Facilitated Disinformation (TF-DI)

Disinformation is content that is false, with the intent to deceive people and cause harm. Despite the central role of technology companies in upholding human rights, preventing the spread of disinformation, and protecting the public from its harm, in the new digital media environment, it often increases the pace and broadens the scope of disinformation dissemination. This dissertation refers to this as Technology-Facilitated Disinformation (TF-DI) [12]. The most important difference made between misinformation and disinformation is the concept of intent to harm and criminality. Disinformation has been considered a crime to resolve rather than just a negative social phenomenon. However, in practice, it is difficult and requires large resources to track down all the sources and prove harmful intent. One of the labels given to disinformation spreading practices aiming to influence elections is 'coordinated inauthentic behavior' [5]. The criminality perspective of disinformation is difficult to seek consensus on, but we argue that disinformation itself may not uniquely be categorized as criminal activity classified as cybercrime [13]. However, disinformation techniques and effects can be utilized to achieve specific targeted criminal activity. This dissertation will look at the intentional criminal offense of utilizing TF-DI to commit crimes in human trafficking, especially recruiting victims via online job advertisements.

The drive of disinformation is multifold, first adapting the two cybercrime categories: cyber-enabled and cyber-dependent crime [13]. Cyber-enabled crime involves traditional offenses that have moved online, such as fraud and theft, occurring in both physical and cyberspace. Cyber-dependent crime of disinformation involves offenses that happen solely because technology made them possible. In the case of disinformation, examples include unprecedented widespread disinformation campaigns targeting elections and algorithm manipulation, such as echo-chamber effects. This means when individuals interact with online content that only aligns with their existing views through social media actions, they reconfirm their beliefs, leading to the persistence of inaccurate beliefs [14].

Apart from the intent and criminality aspects of disinformation, the drives for its perpetuation and stubborn existence also share similar factors with misinformation. Human-centric and technology-centric factors of spreading disinformation can cause it to multiply and easily deceive the general public.

2.3.2 Current Countermeasures to TF-DI

The prominent techniques for combating TF-DI focus on detection, understanding the strategies of malicious actors intending to utilize disinformation, and understanding the affected actors, such as its effects on institutions and individuals.

By reviewing research on TF-DI detection, where Chai et al. (2024) proposes a method that integrates Bayesian topic models and the attention mechanism into a novel 'wide and deep

learning' framework. This addresses the weaknesses in traditional methods where machine learning offers interpretability but lacks depth in representation, and deep learning (DL) models provide a comprehensive range of representation but lack interpretability. They combine both models to enhance the model's ability to analyze textual data with both semantic and syntactic representations. This approach provides overall higher performance and increases effectiveness and interpretability. This model outperforms existing baseline models in capturing disinformation cues like salient topics and words [15].

Another perspective on countermeasures to TF-DI is proposed by Ahmad et al. (2019). The study highlights the importance of understanding attackers' strategic goals and leveraging disinformation to create confusion and operational setbacks for them. They examine Advanced Persistent Threats (APTs) and strategically motivated APTs (S-APTs). Traditional APTs are "malicious, organized, highly sophisticated cyber campaigns aimed at IT networks of specific organizations for long-term access to either obtain information or sabotage operations." S-APTs are "entities that engage in a malicious, organized, and highly sophisticated long-term or reiterated network intrusion and exploitation operation to obtain information from a target organization, sabotage its operations, or both, in support of that or another entity's broader strategic agenda to gain or maintain wealth or power." S-APTs differ in that their objectives are derived from broader strategic goals of third parties such as nation-states or criminal syndicates. They outline stages of APT operations and leverage situation awareness theory to demonstrate how disinformation can undermine the decision-making processes of S-APT operators and their backing entities. By employing techniques such as honeypots and honeytokens, defenders can deceive attackers into believing they have gained access to sensitive and valuable information. This can waste the attackers' resources, gain further intelligence on their tactics, and cause confusion and operational setbacks [16].

Another study dives into the 'Echo Chamber' phenomenon and brings a fresh perspective. Garett (2017) proposed that the real issue is not the fragmentation of the audience into echo chambers but the deliberate spread of disinformation. He points out that the human tendency to seek out attitude-congruent information is not as strong as we would think. Therefore, he urges efforts to address disinformation campaigns rather than worrying about people being segmented by 'echo chambers' [14]. By focusing on understanding how disinformation is spread and its effects, the paper by Davis and Beck (2023) identifies the digital materiality elements: modular content, content flow, and manifold network structures that enable the propagation of disinformation on social media. Along with the three key disinformation affordances: crafting, amplifying, and partitioning, they argue that disinformation disrupts institutional issue fields by altering power centralization, subfield structures, and institutional infrastructure. The authors propose counteractions such as redesigning social media algorithms to prioritize chronological content ranking over engagement-driven content to reduce the virality of disinformation, implementing policies and practices to remove disinformation, and addressing information literacy by providing education to users to make them more vigilant regarding the truthfulness of online content [17].

2.4 Misinformation Risks Posed by Generative Models

In the era of social networks, 5.07 billion users [18] are active on social media, contrasting with traditional news distribution by only a few news outlets. Every user can act as their own news broadcast agent, increasing the challenge for information verification and leading to an unchecked balance in media and information bias. Now, as we step into the era of artificial

intelligence, advancements in generative models, such as large language models supported by pre-trained networks, transformer algorithms, and vast datasets, enable the generation of high-quality content based on users' prompts. Danni Xu et al. refer to such content as Artificial Intelligence Generated Content (AIGC). AIGC is not limited to text form; multiple models allow the generation of AIGC in text, images, videos, audio, and multi-modality generation. The authors examined AIGC in misinformation generation and classified it into three categories: transformation, tampering, and generation [7].

Transformation refers to 'transferring information from one modality to another.' The expansive capability of generative models allows various modality transformations, such as text-based prompts transformed into images, videos, or audio, and captioning of videos and images. The authors identify two primary uses of generative models in facilitating misinformation: 1. Generating assisting/supporting videos or images to back up false information. 2. Creating false misinformation by using pre-existing evidence and materials. Tampering involves 'editing existing contents.' Due to the nature of large language models, only modification prompts are required to alter existing content. Generation involves generating new misinformation content and encompasses random and specific misinformation generation. 1. Random generation is creating attention-seeking content based on popular posts to gain public attention, which can cause distraction and information disorder. 2. Specific generation refers to content generated specifically to fulfill a certain purpose. In the context of misinformation, generative models often suffer from hallucinations, meaning that sometimes a large language model generates false outputs that are decoded wrongly by the transformer and not based on the training data [19]. Consequently, the content generated by large language models is not always truthful, which leads to misinformation.

2.5 Disinformation Risks Posed by Generative Models

On the other hand, in light of advancements made in generative models, the increasing threats of disinformation have been exacerbated, specifically by Large Language Models (LLMs). Barman et al. concur that the capability of these generative models to create highly convincing and context-aware disinformation across various modalities—text, image, audio, and video—can be exploited to automate and scale up disinformation campaigns. Therefore, there is an urgent need for robust ethical guidelines and enhanced interventions to understand the strategies and techniques used by attackers to counter these threats. The authors also propose leveraging these generative models like LLMs with human oversight to assess and combat disinformation [20].

2.6 Case studies

2.6.1 Technology Facilitated Human Trafficking

Human trafficking is a serious global issue, generating 150 billion dollars in 'profits' and affecting about 25 million people worldwide. [28] It involves the exploitation of individuals through force, fraud, or coercion for various forms of labor or commercial sex. This complex issue is deeply entrenched in various legitimate industries and systems. The effects of internet innovation further facilitate human trafficking activities (technologies facilitate the trafficking in human beings - THB) [29]. The rapid exchange of information and communication exposes more victims to traffickers. This includes the use of online recruitment by posting false job

Commented [JH2]: How does technology facilitated HT: look at indicators notes

advertisements or establishing online relationships to lure victims; operations online by advertising 'sales' of victims and establishing online business fronts; and controlling victims through online platforms by monitoring them, sending threats, or blackmailing with sensitive images even after they have left the trafficking situation [21]. Another characteristic of online technology is anonymity, such as the use of dynamic IP, multihoming, pseudonyms, and registering websites offshore to obscure identities [30]. This makes traffickers and THB activities harder to trace and detect.
This shift from traditional physical spaces to the online virtual space has significant impacts on activities linked to THB [31]. It affects the entire trafficking chain, from identifying potential victims, recruitment, and transportation to harboring the victims and exploitation [31].

2.6.2 Current Countermeasures to Technology Facilitated Trafficking of Human Being (TF-THB)

The advancement in generative AI tools may allow offenders to leverage technology to further develop their human trafficking campaigns, but these tools can also be used to combat human trafficking. Applications include mapping networks and revealing criminal groups, tracking financial transactions from trafficking activities, and conducting geospatial analysis to predict illicit activities such as human trafficking [23]. In the context of this dissertation, we specifically examine how TF-MI and TF-DI enable human trafficking as a criminal tactic or exploit its effects, and the associated countermeasures.

The detection of MI and DI markers is central, and this is also true for detecting human trafficking MI and DI content. Brewster et al. (2014) propose utilizing open-source data from the web and social media to enhance the detection and analysis of human trafficking indicators. The proposed framework integrates strategic and operational indicators from credible sources, materials from anti-trafficking activities, and social media content, using NLP and big data analytics. The aim is to improve situational awareness and decision-making for law enforcement agencies (LEAs), which is a valuable proposal and an incremental part of detecting MI and DI related to human trafficking offenses [24].

Another detection study by Vajiac et al. (2023) developed DeltaShield, utilizing information theory. This tool is designed to identify DI markers indicating human trafficking—micro-clusters of near-duplicate documents. This method highlights common phrases, automatically detects variable slots within clusters, and generalizes. It is also applicable to various languages and domains, such as Twitter bot detection. DeltaShield achieved 84% precision in detecting human trafficking ads, making it a practical tool for identifying organized illicit activities through online DI advertisements [25].

Another approach by Kapoor et al. (2017) involves extracting geotags from webpages associated with human trafficking. The approach combines contextual clues, constraints, and the Geonames knowledge base within an Integer Linear Programming (ILP) model to increase overall performance. This method effectively addresses issues in the human trafficking domain, such as mystified language and unclear terms, by using relational and prior knowledge, significantly improving the identification and tracking of online human trafficking activities [26].

Besides detection, one of the countermeasures around TF-THB calls for addressing public awareness regarding accurate information and the complexity of the issue of human trafficking. Erickson and Stoklosa (2022) highlight the role of the QAnon conspiracy movement in distorting public understanding and response to trafficking. They spread false narratives via social media, rooted in extremism, oversimplifying the complex issue of trafficking, and diverting attention from its systemic causes while reinforcing harmful stereotypes. QAnon's disinformation tactics include false storytelling and utilizing social media's echo chambers and filter bubbles to reinforce beliefs among individuals. They recommend that healthcare providers and the broader health sector promote accurate information and inclusive policies to address the root causes of trafficking and support diverse survivor demographics, thereby combating disinformation and supporting trafficking victims [22].

2.6.3 Current Countermeasures in (TF-THB): Detect Fraudulent Job Advertisement

This section sheds light on how disinformation tactics and effects are used for human trafficking-related recruitment by spreading false job advertisements online via multiple channels. One of the shifts in THB activities from physical to online spaces is the recruitment of victims [31]. Various tactics and means of using the internet are employed by traffickers, such as promoting deceitful job advertisements with false promises to the victims, placing travel opportunities on advertisement sites, or through international marriage agencies or dating platforms. Additionally, traffickers make personal connections with the victims by initiating contact via online chatrooms or mainstream media. This dissertation focuses on how traffickers use deceitful strategies through online channels to spread disinformation through job advertisements and information to lure/recruit potential victims [30].

In Volodko et al. (2019), the authors investigate the prevalence and common indicators associated with labor trafficking advertisements online, sourced from a Lithuanian job-seeking website. They analyze 430 online advertisements to determine whether these ads contain indicators of labor trafficking and to explore the patterns or correlations between the indicators. Utilizing quantitative content analysis, the indicators are coded in descriptive format (sourced from research contents), and descriptive statistics, Poisson regression, and chi-square tests are used to examine relationships between the advertisements' characteristics. The authors found that an overwhelming number—up to 98.4%—contained at least one indicator, with 63.7% of ads containing two to four indicators, suggesting that the indicators may not allow for effective discrimination of HT-related job ad content. Additionally, the authors discovered that violations of minimum wages and working hours are possibly the most salient indicators of intended exploitation. Destination countries like Cyprus and the Netherlands had higher counts of wage violations, particularly in industries like hospitality, forestry, and horticulture [32].

The recruitment tactics used by traffickers, as identified in the Mekong Club report, include: 1. Creation of fake recruitment campaigns, 2. Private job matching groups, 3. Through personal networks, 4. Posing as professional clients (e.g., customers inviting magicians, interior designers, and tour guides to visit Cambodia and provide services), and 5. Kidnapping from the street. Tactics 1 to 4 can be performed via the use of the internet.[33]

Janušauskienė (2013) adopts multiple methodologies, including documentary analysis, interviews, and observations. The results confirm the prevalence of false job advertisements in labor trafficking, with many cases originating from online job offers. This creates a grey

area in labor trafficking, as the deceitful nature of online recruitment leads victims to not realize they are victims of trafficking, further complicating the identification and protection procedures for potential victims. The authors also concur that the lack of awareness and expertise among frontline officials hinders the identification of victims, which calls for prevention efforts from public education systems and employment agencies, especially through proper controls around online employment job boards.[34]

Recent research by Zhou et al. (2024) involved collecting over a quarter-million job postings from eight online job boards targeting Chinese-speaking immigrant job seekers from the period of 2006 to 2024. The categorization process was based on self-reported information from users, and the investigation focused on the types of advertised opportunities, preferred modes of contact, and the frequency of postings to uncover patterns of suspicious ad postings. Additionally, using temporal analysis, the study explored how external events may influence traffickers' increased targeting of vulnerable victims. The results show patterns such as the use of short, detail-lacking ads and hotel locations for interviews, especially contact numbers previously listed on escort service websites, all flagged as potential sex trafficking hazards. The use of phones as contact points can amplify the risks of harm due to the difficulty in tracing them. The study also pinpointed industries prone to suspicious activities (e.g., massage parlors, nanny services, restaurants, and driving). By focusing on culture-specific platforms (Chinese-speaking job seekers) and encouraging the monitoring of contact details to track trafficking activities, the study also noted that external events (e.g., Covid-19, the Ukraine-Russia War, the Israel-Hamas War) correlate with a spike in ad posting frequency, giving traffickers more opportunities to prey on vulnerable demographics [35].

2.6.4 Methodological challenges around HT research

The methodological challenges around HT research are a significant concern. They extend beyond technological restraints; in fact, human trafficking issues originate from broader, complex socio-economic and cultural causes. These complexities present significant challenges in developing protection, detection, and prevention mechanisms from empirical and actionable technological perspectives. **[30]**

Laczko (2005) points out several challenges; one of the most challenging problems for researchers is the ‘hidden population’ phenomenon (victims/survivors of sexual trafficking, traffickers, or illegal migrants). It refers to ‘a group of individuals for whom the size and boundaries are unknown, and for whom no sampling frame exists’ [27]. Researchers face difficulties in obtaining a truly representative sample of the population, resulting in a significant lack of reliable data. National and global figures are often no more than ‘guesstimates’, and research often relies on a small sample of data. The reasons for this are often because trafficking is a ‘clandestine’ activity, and many cases go unreported, or because the victims are reluctant to go to the authorities due to their involvement in stigmatized and illegal activities, such as their illegal status in destination countries or a lack of communication means (trafficked victims often do not speak the language of the destination country). [36] Additionally, the inconsistency of definitions around human trafficking poses further difficulties [32].

Tyldum and Brunovskis (2005) also criticized the highly politicized policies and issues related to human trafficking, such as ‘prostitution, labor market protection, and immigration laws’. Key actors with specific agendas can influence how they publish or use the data. The authors provide a structure of the stages of trafficking, advising a comparative method in investigating the stages a victim undergoes during trafficking and in relation to the process.

Furthermore, they emphasize the importance of potential variances within different stages, which can lead to different research approaches and methodologies.[27]

This dissertation focuses on researching the initial stage where people are at risk of being trafficked, associated HT criminal acts of recruitment of victims, specifically on online platforms via job advertisements, with a narrow focus on labor trafficking. The methodological challenge stems from the root issues of identifying forced labor. First of all, the 'hidden population' phenomenon makes it difficult for victims subjected to forced labor to report themselves to the authorities; often, they fear the risks of being deported, and traffickers tend to prey on victims from lower socio-economic statuses and underdeveloped countries. [30] Therefore, they may also be reluctant to share personal experiences. Another ethical issue arises: what if victims ask the researcher/interviewer for help? Secondly, there is bias in official sources and policies. A case where a German couple deceived and severely exploited eight young women, resulting in one woman committing suicide in 2003, was not counted as a trafficking case due to German law at the time only considering trafficking for sexual exploitation. This example shows bias in their perception of trafficking as linked to women in sexual exploitation, a similar conception found in mainstream media. [37] This may result in overlooking labor trafficking issues, exacerbating the lack of data, research, consistent methodologies, and research around this issue, in addition to the lack of clarity around HT-related definitions. Further literature sheds light on the distinction between "non-trafficked victims of forced labor" as opposed to "trafficked victims of forced labor," where recruitment mechanisms into forced labor are linked to the labor market in the destination country, making it more difficult to identify the victims and therefore allocate resources for protection, prevention, and aid. [37]

Chapter 3 Methodology

The research methodology for this dissertation adopts a top-down quantitative approach, which is the most fitting approach. Figures 3 elaborates on the research process and methodology. After identifying a real-world problem and research gap, our aim and objective is to devise and advise a solution. An essential part of the research process is utilizing a fitting research methodology, which gives scientific validity and allows fellow researchers to follow up on the work. It enhances scientific integrity by allowing others to repeat the scientific experiments that have been conducted.

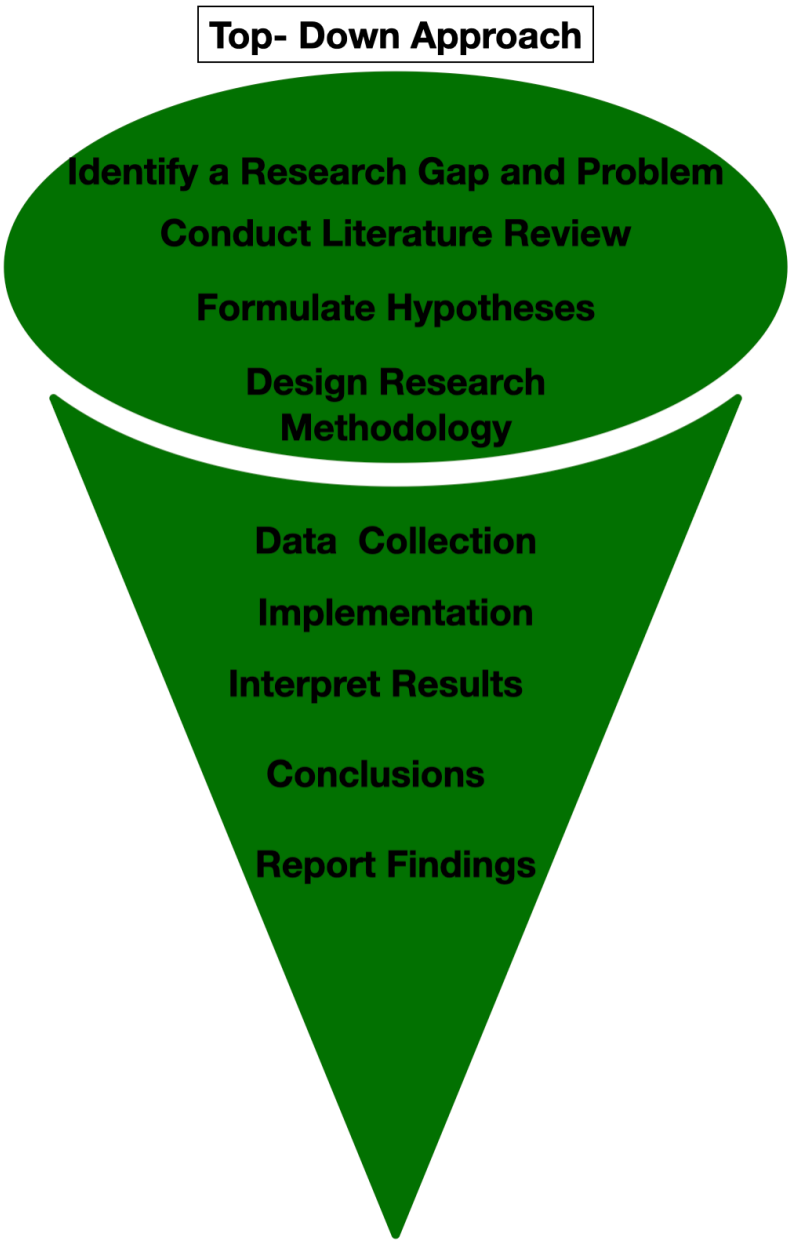
3.1 Research Methodology

Overall, we conduct a literature review regarding misinformation and disinformation, focusing on technology-facilitated misinformation and disinformation challenges. We then identify some of the state-of-the-art machine learning techniques and solutions proposed by academics. Next, we examine how organized crime can use technology-facilitated misinformation and disinformation to manipulate and commit human trafficking crimes. We then identify and design machine learning applications to combat this challenge. The model will be trained and tested using selected datasets. Lastly, the models will be compared, their limitations assessed, and suggestions for future research will be made.

3.2 Implementation Methodology

In this dissertation, we will use Python on Google Colab to implement a machine learning solution, incorporating natural language processing techniques, to detect disinformation with relevant markers that reflect human trafficking criminal activities. We aim to train the model on the dataset and able to detect human trafficking ads.

Figure 3.0



Commented [JH3]: Figure2

Chapter 4 Implementation

The chapter discusses the process of sampling data used in this dissertation, particularly the vocabularies collected that contain Human Trafficking-related Job Ads indicators. This is followed by the implementation of two Machine Learning Classifiers—Logistic Regression (LR) and Support Vector Machine (SVM)—and the analytical procedures. The coding and datasets have also been made available via:

The data are generated using the data generation tool generatedata.com, while the vocabularies are collected in a separate process. The datasets used for model training contain two classes and are divided into training and testing datasets.

- Training dataset:

Class 1 (Human Trafficking indicated job ads) = 1000

Class 0 (Legal Job ads) = 1000

- Testing dataset (unlabeled):

Class 1 (Human Trafficking indicated job ads) = 1200

Class 0 (Legal Job ads) = 1200

4.1 Sampling and Data

To obtain the datasets used to train the models, the following options were attempted:

1. Conducting detailed literature reviews on the relevant subject and identify researchers to contact and inquiry sharing of their datasets.
2. To scrap data online.
3. Identify thematic categories via literature reviews and select keywords/phrases to compile a dictionary containing two classes (0- legitimate Job ads / 1- Human Trafficking related job ads)

The key challenges in the methods above are the unavailability of the researchers to share datasets, and scraping data online is costly for time and resources. Additionally, the nature of the human trafficking related topic requires a consultant in legal department or experts to double check and confirmed the scrapped data positively used for human trafficking organizations. Considering the time constraint and resources, the option three is chosen.

Option three allows the preservation of the scope and explainability of the datasets, by compiling the human trafficking (HT) and non-human trafficking job ads dictionaries with supervision and ensuring the HT indicators presence or excluded in respective classes of the datasets.

The process of sampling data as follows: first pinpoint the problem, to detect HT job ads, then conduct literature reviews to identify eight thematic categories relate to the HT job ads and select key words represent those eight categories respectively to compile the dictionary.

4.1.1 Literature Review on Human Trafficking Detection Indicators

To identify the indicators commonly presented in Human Trafficking-related online job ads, it is crucial to understand the nature surrounding the criminal acts of human trafficking. There have been various attempts to research and seek consensus on the terminology and definitions related to human trafficking criminal acts. To begin with, the definition provided by the UN's 2000 Protocol to Prevent, Suppress and Punish Trafficking in Persons, Especially Women and Children, supplementing the United Nations Convention against Transnational Organized Crime, in Article 3 - use of terms, 'Trafficking in persons' means [38]:

1. The *criminal acts*: recruitment, transportation, transfer, harbouring or receipt of persons.

2. The *means* being used: (of threat or use of force, coercion, abduction, fraud, deception, and the abuse of power or of a position of vulnerability)

3. The *purpose* of trafficking another person or gain control over another person (sexual exploitation, forced labour or servers, slavery (include servitude or the removal of organs).

The key in this definition are the criminal acts of recruitment of victims and the means being used (deception). The purpose of trafficking will be overlooked in this chapter, due to the vagueness around definition of *purpose*- criminal intents to traffic people , which is difficult to quantify the problem. [37] Therefore, the implementation of machine learning detection, strictly concentrate on the detection of disinformation strategy using to deceit, using online job advertisement to lure victims to be trafficked into forced labour and other forms of exploitation.

Another definition identified three stages in human trafficking by Tyldum and Brunovskis: [27]

- People at Risks being trafficked

- Current Victim of trafficking

- Former Victim of trafficking

The dissertation aims to prevent the first stage: *people at risk of being trafficked*, by understanding and detecting the deceptive strategies used to deceive and recruit victims. Despite the detailed logistics and the process of recruitment remaining unclear due to the constraints around the methodological challenges in HT research (see chapter two, subsection 2.6.4).

Through a detailed literature review, six indicators commonly present in HT job ads are identified: **Accommodation Provided**; **Transport to Destined Country**; **Job Conditions (salary, working hours)**; **Guaranteed Visa/Stay in the destined country**; **Initial Settling**; and **Victim Descriptions (represented as job requirements).** In addition to these indicators, four categories were added to construct a complete job ad: **Job Industries**; **Description of 'Company' & Employment**; **Job Descriptions (responsibilities).**

Commented [JH4]: Designing Trafficking Research from a Labour Market Perspective/ The ILO Experience

Commented [JH5]: Describing the Unobserved/ Methodological Challenges in Empirical Studies on Human Trafficking

Table 1 shows the five primary indicators that make up the vocabulary for HT job ads and the relevant definitions/findings from the literature, along with the sources. The likelihood of a job ad presenting a combination of these five indicators simultaneously being related to trafficking activities is higher. Indicators are more useful in combination than alone, according to Volodko et al. [32]. From the research findings, Accommodation Provided is the most commonly presented indicator in suspicious HT job ads, but its single presence's validity is unclear. Attractive Conditions (presented as Wages/benefits) [34] related indicators tend to be higher for job recruiting. However, indicators such as Accommodation, Transportation to the Destined Country, and Help with Initial Settling are ambiguous to interpret in determining whether an HT job ad would definitely present these three indicators individually.

Table 1.0

Indicators by categories adopted in the vocabulary:	Definition and basis from the literatures	Sources
1. Accommodation Provided	*"In many cases, victims are forced to repay the cost of their transportation and accommodation to their exploiters, and are kept in debt bondage for indefinite periods."*	Europol 2016, p.22
	Ads are full of unrealistic promises, such as.. "free accommodation".	Janušauskienė 2013, p.337
2. Transportation To Destined Country	*Mobility and rotation of victims are key features within this criminal market.*	Europol 2016, p.2, 22
	The traffickers may also assist with transportation, including in some cases the necessary documentation.	UN Human Rights 2023, p.13
3. Help with Initial Settling	*'To conceal their identity, traffickers frequently open and control bank accounts in the victims' names'*	Europol 2016, p.26
4. Victim Descriptions(represents as Job Requirements)	*Most reported victims are male EU nationals originating from Bulgaria, the Czech Republic Estonia, Poland, Romania, and Slovakia.*	Europol 2016, p4
	Western and Southern European Member States are key destination countries within the EU. (Free movement)	Volodko et.al 2019, p.5
	Among member states, more registered victims per capita came from Romania, Bulgaria, Poland, Slovakia and Lithuania (2010–2012)	Volodko et.al 2019, p.8
	80% labour trafficking victims identified from EU member states (2009-2014)	Volodko et.al 2019, p.8
	trafficking victims tend to have limited knowledge of the local language of the destination country	Volodko et.al 2019, p.14 (footnotes)
	"Most of the victims we're seeing are from underdeveloped countries,"	Latonero 2011 p.17
	Many victims being deceived often come from Southeast Asia, including Malaysians and Thais, but also from China, Taiwan, Brazil, the United States, Ethiopia, and other countries.	Clack et.al 2024, p.7
	By means of systematic irregular deductions from wages… removal of passports, and threats…, combined with their often poor language skills and a lack of awareness of their rights	Craig et.al 2007. P.37
5. Attractive Conditions, benefits: (salary, perks)	*Job ads are usually placed in the most attractive way in order to attract more potential employees.*	Janušauskienė 2013, p.337
	…the promise of high wages for a relatively simple job, free housing and transportation to the country in question and limited requirements of the candidates .	Volodko et.al 2019, p.11

Commented [JH6]: Table 1

Consider the possibility that Eastern Europe is more dependent on cheap migrant labour [32]; it is reasonable for legitimate companies to set up additional help for migrant workers. Nonetheless, this creates a dependent relationship between the workers and employers, increasing the risk of trafficking. Therefore, the combination of the three indicators with the other two raises reasonable doubt about the legality of a job ad. Additionally, low job requirements attract more vulnerable demographics, which, according to research findings, often include victims from poorer socio-economic environments. These factors, combined with a lack of educational awareness about the job market, increase the risks of trafficking. The indicators of Victim Description (represented as job requirements) are presented as relatively low requisites for a job. Therefore, the vocabulary for HT job ads necessitates that these five indicators be present. The vocabulary also includes an additional category: Initial Settling, wherein keywords found include additional information such as opening a bank account, taking care of necessary documents, or taking the victim to the workplace. These are meant to complement the five indicators, rather than to be a specific scope of content.

In addition, two categories are added to the vocabulary to ensure the completion of the job ad structure: Job Industries; Description of 'Company' & Employment; Job Descriptions (responsibilities). The literature shows that suspicious job ads commonly advertise low-skill manual labour job industries, such as construction, food production, manufacturing, hospitality, and non-food-related packaging [32]. However, additional literature suggests that glamorous job industries, such as modelling or marketing positions, also record victims being deceived into forced labour and other forms of exploitation [39]. The Job Descriptions category complements indicators of Victim Descriptions, where they indicate a low bar to apply [32]. Similarly, the Description of 'Company' & Employment complements the structure of job ads, though it does not necessarily contain set keywords, as no concrete evidence is found that the kinds of descriptions of a company can indicate connections to human trafficking organizations.

4.1.2 Selection of Key Words

By identifying the five indicators and three complementary categories, the search for plausible keywords representing these indicators was conducted. The search for keywords primarily involved a few samples of confirmed human trafficking ads from research papers and news outlets. Table 2 and Table 3 present the findings and examples. However, despite exhausting the online search, only 12 ads were identified, which is not significant enough to compile a vocabulary to use in generating a data tool.

In the research by Volodko et al., the authors selected the source: darbasuzsienyje.org (translated: workabroad.org), which is a Lithuanian job board for workers seeking work abroad. Although the authors noted that it is not confirmed whether any of the advertisements used in the research were indeed linked to human trafficking activities, the keywords and sentences can still be selected for this dissertation. A total of 300 advertisements were viewed and translated from Lithuanian to English using Google Translate. Out of 300, 59 were selected based on manual observation where they contained more than three indicators in an ad, and keywords from the job ads were then categorized into respective categories.

4.1.3 Human Trafficking related Job Advertisements Dataset Vocabulary (By categories)

In the end, two separate vocabularies were compiled for the training dataset and testing dataset for class 1 - Human trafficking-related job ads. In the training dataset, 30 rows of keywords and phrases were identified for each category. In the testing dataset, 40 rows of keywords and phrases were identified for each category.

4.1.4 Legal Job Dataset Vocabulary (By categories)

For class 0 - Legal job advertisements, the categories were constructed based on observations of confirmed legal job advertisements via LinkedIn. Six categories were identified: Description & Requirements (Who we are & Job Summary); Job Summary; Key Responsibilities of the Job; Job Requirements (Eligibility, Availability, Skills and Experiences); Preferred Qualifications but not Essential; and Benefits and Perks. Similarly, two separate vocabularies containing keywords and phrases representing the respective categories were compiled for the training and testing datasets. Additionally, 30 rows and 40 rows of keywords and phrases were individually identified for the training and testing datasets.

Table 2.0

Sources:	Human Trafficking Job Ads Scam (confirmed)			
"Spotting the signs" of trafficking recruitment online/ exploring the characteristics of advertisements targeted at migrant job-seekers	England - LOGISTICS warehouses ([SUPERMARKET BRAND], [SUPERMARKET BRAND] - packaging jobs) [Recruitment agency name] - Guaranteed employed 10 years running !! Great Britain Job Description GUARANTEED EMPLOYEMENT !! We are one of the oldest employment agencies in Lithuania and we continuously offer permanent, legal and well-paid jobs in factories, warehouses, manufactory, printing houses and hotels in England. WE ARE URGENTLY OFFERING PERMANENT, LEGAL AND WELL-PAID JOBS IN LONDON; "Jobs in logistic ("[SUPERMARKET BRAND]", "[SUPERMARKET BRAND]", "[SUPERMARKET BRAND]" and other) warehouses (toy, food product, clothing packaging, loading, "order picker" positions): Hourly pay; Up to 60 working hours per week; Wage on average 300-500 GBP (pounds) per week/ 7.83 - 11 GBP (pounds) per hour; Wage paid on time (every week) !! All social benefits !! Jobs for both men and women; English language knowledge is not necessary! Experience not necessary ! Good living conditions! We take you, greet you, provide lodging for you and take care of the necessary paperwork. We have 8 spots !! Leaving on 3, 10, 17, 24, 31 August, in September and October !! YOU CAN FIND MORE JOB OFFERS on our website [WEBSITE NAME], calling us or visiting our company headquarters (FREE consultations). [website name] Tel.: [telephone number]; [telephone number] E-mail: [e-mail] --- Tax returns for those who worked abroad (England, USA, Norway, Ireland, Holland, Germany, Australia, Canada) !! --- Additional information GUARANTEED EMPLOYMENT. We are the oldest and leading employment agency in Lithuania and we offer well-paid, permanent and legal jobs in warehouses, factories, plants, manufactories in England. Requirements Diligence, honesty, integrity. Wage 1800			
LITHUANIAN MIGRANTS AS VICTIMS OF HUMAN TRAFFICKING FOR FORCED LABOUR AND LABOUR EXPLOITATION ABROAD (Advertisement 2)	NORWAY- CHEESE FACTORY Wages 135nor/krone17 Euro per hour. NO NEED TO KNOW LANGUAGE. You will work near OSLO and TRONDHEIM. WORK IN HOT OR COLD ENVIRONMENT!!! Cheese production line, work in 3 shifts (morning, afternoon, and night). Specific work: at conveyer, at a production line, packaging line, loading, cleaning of production premises. NO NEED TO KNOW LANGUAGE. You will work near OSLO and TRONDHEIM. Contract for 5 months. Transport and accommodation - 400 kron per week. Departure in JUNE-JULY. Already registering			
LITHUANIAN MIGRANTS AS VICTIMS OF HUMAN TRAFFICKING FOR FORCED LABOUR AND LABOUR EXPLOITATION ABROAD (Advertisement 3)	Looking for young workers in factory in Netherlands. The job is to package smoked fish products Departure this or next week. No need to know foreign language. Job starts immediately after arrival. Arrival on weekend, starting the job on Monday. Rent of living premises deducted from wages, no need to pay upon arrival. You need to have some money for food only.			
Stuck in Trafficking: A Contextual Analysis of Human Trafficking	*"DAILY CASH PAYOUTS $500-1200 daily Safe professional clients -- WE SCREEN ALL. Upscale locations All bookings and advertising provided A decade of experience Innovative and organized management WHO WE EXPECT TO WORK WITH: Attractive females 27-45 YRS Sexy and open-minded and personality Must be reliable and trustworthy Must have a reliable car Must be NOT OVERWEIGHT Must always remember that CLIENT SATISFACTION IS NUMBER ONE Priority SEND 1 PHOTO AND CONTACT INFO WHEN REPLYING!! WE WILL CONACT YOU IF INTERESTED WITHIN 24 HRS."*	Start Modeling Career with NEW Lingerie Magazine! - 34 Our NEW Victoria Secret style lingerie magazine is looking for some sexy models that are comfortable with their body and are interested in being featured in this new upcoming magazine. MODELS MUST BE WILLING TO TRAVEL! Models must be 18 or older. BEGINNER MODELS WELCOME! NATIONAL EXPOSURE! HOUSING INCLUDED! If you are interested in starting a REAL MODELING CAREER send PHOTOS and PHONE NUMBER to this email. We will develop you! We will build website and build your professional modeling portfolio! Also feel free to call us at 503 268 2446. Thank you. (PHOTOS AND PHONE NUMBER)	*"Adult Entertainers: Make Today! We're looking for new faces. attractive Females escorts Wanted for up-Scale Adult Establishment. Please tell us about yourself and send place and body pics to e-mail: theligneriestudiotampa@gmail.com". Now Hiring Lingerie Models for 6PM-2AM shifts and weekends.Start Today!! Text 813- 540-4094 - 29 Remain independent and make Daily Clean Up-Scale Environment On Facebook: https:// www.facebook.com/theinvolditampa/ ATM available inside In-call Only Were currently Hiring please inquire within. The Studio 1053 West Busch Blvd Tampa FL 33612 New Phone # 813-933-9411 Hours: Sun-Tue 10 AM-2AM Wed-Sat: open 24 Hours Look For The Red Lights Brandon clearwater st Pete New Tampa tampa"*	*"Ladies ONLY, Looking to become the next BIG ADULT SUPER STAR? Come join our ELITE talent management company! Have fun and be safe STARRING in NON DEGRADING HIGH END Adult Film Magazines ie. Penthouse Hustler Cheri etc. $800+/per shoot. STEADY, LONG TERM, QUICK, AND EASY MONEY! PLUS FREE HOUSING!! SAVE $200/MK THAT THE OTHER AGENTS CHARGE AND YOU WON'T EVEN GET YOUR OWN ROOM OR BATHROOM. WE CAN ACCOMODATE THAT TOO! WE HAVE 2 OTHER MODEL HOUSES THAT YOU WILL NOT HAVE TO SHARE YOUR ROOM OR BATHROOM FOR THE SAME FEE. PLENTY OF REFERENCES, WE ARE HAPPY TO PUT YOU ON THE PHONE WITH OUR CURRENT MODELS. Just text some smiling selfies and bikini pics, or give a call to Company owner Bella. If email is easier for you, you can just email pictures to her"*

Commented [JH7]: Table 2

Table 3.0

https://www.propublica.org/article/human-traffickers-force-victims-into-cyberscamming	Recruitment news_Customer Service staff both genders for 10 positions. Basic Salary: 1000 + Commission Working Hours: 3:30pm-3:30am Age 18-29 years old (boy/girl) Need to know English 50% up + use computer Typing 20-30 words/ minute No experience ok ok some experience is also good 1 day off/ month (work 6 months can rest 15 days) Provide accommodation and 3 meals Location GM Kampong Som (Ou Tte)			
From Fake Job Ads to Human Trafficking The Horrifying Reality of the Human Trafficking Scam Trade	General Assistant Job Requirements: Age: 21-29 years old Gender: female, no nationality (foreigner is preferred) Education: college degree or above (1-2 years after graduation is preferred) Behavior and conversation Fluent in English, smart, smart, good image and temperament Strong stress resistance, smooth communication with people, optimistic and cheerful personality, and not sloppy in doing things. Responsibility description: Assist leaders in related business activities and daily itinerary coordination; Daily entourage translation and communication Work arrangements; Assist executives in daily management of vehicles, nannies, and bodyguards Salary range: basic salary 15k-25k RMB, specific negotiable; Working hours: 9 hours, 1 day off per week; [7 Benefits) (All positions in the company are entitled to) [Employee accommodation) The company provides accommodation, hotel-style apartments, and single-person accommodation in high-end residential areas with complete facilities, including: WIFI, gym, swimming pool, water heater, gas furnace, air conditioner, etc., so the position 24-hour standby is required, and confidentiality is strong, so it is mandatory to prohibit overnight stays. [Paid annual leave] If you voluntarily give up paid annual leave, you can apply for an annual leave subsidy; a one-time reward of 7,000 yuan (as the level increases); [Birthday benefits) The company will issue a birthday gift of 5,000 pesos in the month of the employee's birthday; [Holiday Benefits) During the Spring Festival, Dragon Boat Festival, and Mid-Autumn Festival, in-service employees can enjoy this benefit. At that time, the company will distribute this benefit in the form of currency or in kind; the salary of the in-service employee is 3 times during holidays: Dragon Boat Festival, Mid-Autumn Festival, and Labor Day: (More at the end of the year Salary) It depends on the position, and the salary is 13/14/15/16 at the end of the year; Telegram: @HR_Eleven			
https://news.trust.org/item/20210916120210-olp4a	PHNOM PENH JOB OFFER Job description* # Online customer service marketing Job location# PHNOM PENH. Requirements 01. Must know English language proficiency and computer typing skills. (Basic salary#$950+commission 5-10%) Age 18-35 Company providing by Visa, Work permit, Food & Accommodation. First Contact six months. Job vacancy#20 person Including two translator.			
https://www.euronews.com/next/2023/10/25/revealed-global-whatsapp-jobs-scam-worth-100m-linked-to-china-and-cyber-slavery-networks	!! JOB OPPORTUNITY !! Marketing Online Work TIME : 1-2 hours (Flexible hrs) INCOME AGE £100-600 daily :20 years old and above EDUCATION : no qualifications required REQUIREMENT: smartphone or computer with good internet connection (training is provided)	INCOME : £100-600 daily AGE above 20 years old and EDUCATION required : no qualifications REQUIREMENT : smartphone or computer with good internet connection (training is provided) Staff, Fresher, Students, Housewives, anyone can work Staff, Fresher. Students. Only serious applicants are needed. You can contact us at WhatsApp:	Nous recherchons un télétravailleur pour des hôtels et des attractions touristiques internationales, pouvant travailler à distance 30-50 minutes par jour, 22+ ans, esprit clair, travail facile, évaluation et réservation uniquement, 100-700€ par jour, payé le même jour, Whatsapp Employeur :33644639766 We are looking for a teleworker for hotels and international tourist attractions, able to work remotely 30-50 minutes per day, 22+ years old, clear mind, easy working, assessment and booking only, 100-700€ per day, paid the same day, Whatsapp Employer: +33644639766	Vacant Positions (Full/Part Time) Position Requirements: Swiss citizen, PR or visa holder (general terms and conditions) Age 20 and above No experience necessary Work 60-180 minutes per day Training provided Payment in USDT or CHF Basic pay from $5500-$9000, CHF 5000-8500 (not including commission)
https://www.snopes.com/fact-check/interview-for-trafficking/	Dear Ruby, Due to our company's continued growth and expansion, we currently have openings for summer work. We would like to invite you to apply. Vector is an international company established in 1981. We are currently expanding locally and have summer customer sales and service positions available. These positions will offer several unique advantages for those students accepted: Full and part time openings with flexible schedules $16.00 base-appt. Training provided - no experience necessary Valuable resumé experience May continue during school in the fall Conditions apply All ages 17 Opportunity to advance We are filling openings for Sandy, Salt Lake City, Northern Etah County, Park City, Tooele, and surrounding areas. schedule, class schedule or previous work experience. You may apply regardless of your major, current work look forward to meeting you and explaining the details of the position. interview, please contact receptionist 601, 747-5210 Monday through Friday To schedule an between 9:00 am and 6:00 ye or	TEENS FOR AFTER SCHOOL JOB EARN $60, $80, $100 PER WEEK!!!! WORK EVENINGS & SATURDAYS FREE TRIPS & ACTIVITIES TRANSPORTATION PROVIDED 14 YEARS AND OLDER BONUSES!! BRING YOUR FRIENDS!!		

Commented [JH8]: Table3

4.2 Data Generation

The vocabularies for the two classes were created and uploaded into generatedata.com, using the random word selection feature and uploading the custom vocabularies. The word ranges are set based on the vocabulary word count; the minimum is set at the median of the word count of a specific vocabulary, and the maximum word count per row is set for the maximum words to be generated. This is to ensure the possibility of covering all the phrases and categories per row during the randomization process. The training dataset generated 1,000 rows per class and 1,200 testing datasets per class. All datasets were generated in CSV format.

- HT_train vocab **max** words in a row: **297**/ HT_test vocab **max** words in a row: **286**

- Job_train vocab **max** words in a row: **157**/ Job_test vocab **max** words in a row: **155**

- HT_train vocab **median** words in a row: **297**/ HT_test vocab **median** words in a row: **286**

- Job_train vocab **median** words in a row: **157**/ Job_test vocab **median** words in a row: **155**

4.3 Model Implementation

Once the datasets are prepared, two machine learning classification models (a baseline model and an advanced model for comparison) are chosen to be trained, and their performances are observed.

4.3.1 Data preparation

A series of text pre-processing steps were undertaken to clean the data and prepare it for training the model. We started by examining our data and conducting Exploratory Data Analysis (EDA) to identify the type of data and visualize the balance of the data. Followed by text pre-processing steps:

1. Encode text to ASCII.
2. Normalize text (lower case).
3. Tokenize the text.
4. Remove Punctuation.
5. Expand Contractions.
6. Remove numbers.
7. Remove special symbols.

Noted, both models will use the same function (clean_text) for text pre-processing to ensure consistency and comparability.

4.3.2 Baseline Model Training: Logistic Regression

The baseline model chosen is Logistic Regression (LR). It is a supervised machine learning algorithm primarily used to discriminate between two classes, making it a good classifier for binary classification [40]. The model works well with small datasets and is efficient [41], which is suitable for the datasets at hand. The LR model utilizes TFIDF feature extraction with the specified remove stop words command to exclude stop words from the datasets, allowing the model to focus on the selected terminologies in the vocabulary. Then, the model uses GridSearch to explore optimal parameters; **C**, **penalty**, and **solver** are chosen to be tuned due to their function in preventing overfitting.

4.3.3 Advanced Model: Support Vector Machine

The advanced model for comparison is the Support Vector Machine (SVM). Similarly, it is a supervised machine learning algorithm. The SVM model aims to find a hyperplane in n-dimensional space that classifies the data points [42]. It is particularly effective for datasets containing multiple features and higher-dimensional problems. Its advantage lies in its kernel,

where a function takes low-dimensional input and transforms it into high-dimensional space; it is also memory efficient. The same feature selection tools, TFIDF and excluding stop words, were applied to the SVM to ensure consistency. Hyperparameters were tuned using Grid Search, with **C**, **Gamma**, and **Kernel** selected for tuning.

Chapter 5 Findings

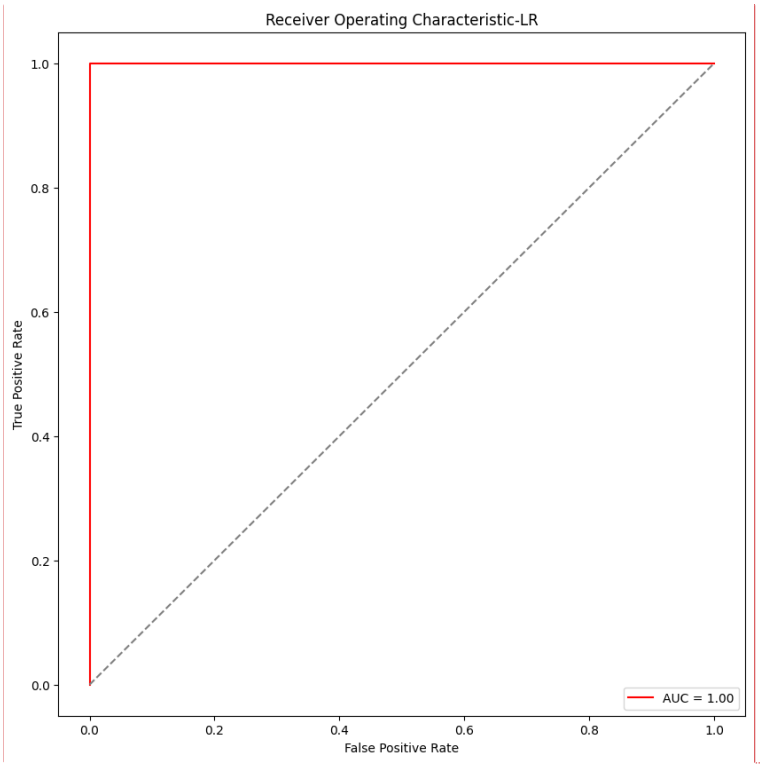
Both models returned 100% training and testing accuracies, despite the changes made in splitting data, tuning hyperparameters, and conducting cross-validation.

5.1 Model Findings and Comparisons

LR obtains 100% accuracy in the training process as well as in predicting labels on the unseen test dataset (not labeled). It correctly predicted 2,400 rows of data into two classes. In comparison, SVM also achieved 100% training accuracy. However, when predicting the unseen test dataset, it misclassified 12 data points into the wrong classes.

Figure 4.0 shows the Receiver Operating Characteristic curve (ROC), which presents the LR model's performance. The X-axis shows the false positive rate, and the Y-axis shows the true positive rate. The curve (red line) rises straight from 0.0 to 1.0, and the Area Under the Curve value is 1.00, all indicating a 'perfect' model.

Figure 4.0

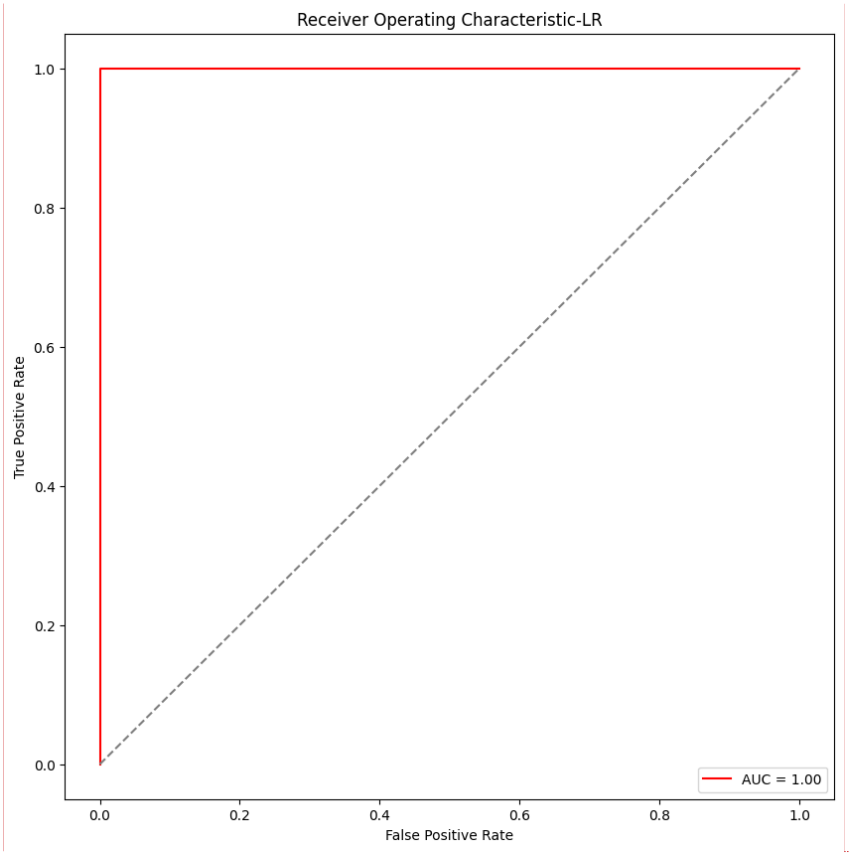


Commented [JH9]: Figure 3

The SVM model's training/testing accuracy is also at 100%. Refer to Figure 5.0, which also shows similar 'perfect' performance. Overall, the models seemingly perform well in

discriminating between training datasets on human trafficking-related disinformation job advertisements and legal job advertisements. However, by incorporating the theories and examining the dataset closely, there are several noticeable observations in addition to the model performances.

Figure 5.0



Commented [JH10]: Figure 4

5.1.1 Model performance on Unseen datasets

Noticeably, the LR model correctly predicted all 2,400 rows of data into the respective classes (1/0), while the SVM model incorrectly predicted only 12 out of 2,400 rows of data.

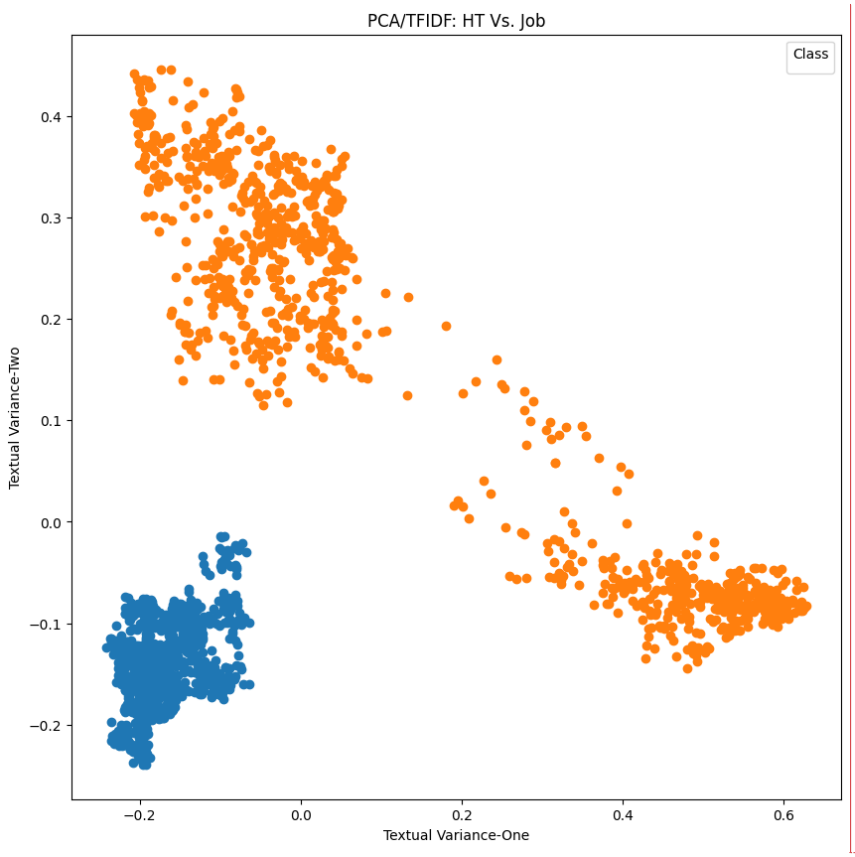
5.2 Dataset Characteristics: Class 1 and Class 0

By performing a Principal Component Analysis (PCA), which reduces the number of TFIDF features to capture the variance in the dataset, refer to figure 6.0, which represents class 1

(human trafficking-related ads) and class 0 (legal job ads). It shows a clear separation of classes. The orange cluster represents class 0 (legal job ads), and the blue cluster represents class 1 (human trafficking-related job ads). The class 1 (blue) cluster is more spread out than the class 0 (orange) cluster. Therefore, the distinct characteristics of the two classes allow models to identify them.

The class 0 (orange) cluster, which represents the legal job ads, is more closely packed and dense, which could suggest that in class 0, the textual data is more harmonized, presenting less variation in the usage of language. In contrast, the class 1 (blue) cluster is more spread out, suggesting greater variance in the language used in class 1. However, they still cluster together in the right-lower bounds of the plot, indicating that there are still some common features in the language used in class 1.

Figure 6.0



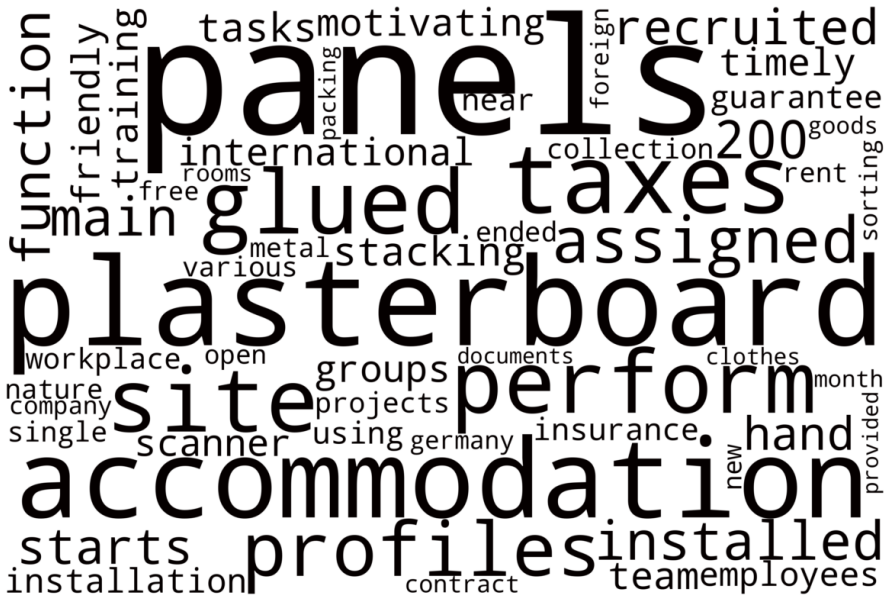
Commented [JH11]: Figure 5

To examine the words more closely, refer to figure 7.0. A word cloud visually represents the most salient words in human trafficking-related job ads; the most frequent words appear larger, while smaller words appear less frequently.

Most salient words (class 1):

- panels /glued /taxes /plasterboard /assigned /site /perform /accommodation /starts /profiles /installed.

Figure 7.0



In comparison to figure 8.0, which is a word cloud representing legal job ads:

Most salient words (class 0):

- calls, fellow, education, potential, informational, raiser, committed, boxes.

By visually examining the salient words in the two classes, and knowing the vocabulary compiled to represent the two classes, words in the class 1 word cloud, such as "accommodation," do fit one of the indicators. However, the rest of the words are not as clearly indicative of human trafficking-related job ads.

Similarly, "education" may represent a higher job requirement in legal job ads, but it is not sufficient to make clear distinctions regarding how certain words represent the respective classes.

Figure 8.0

Commented [JH12]: Figure 7

5.3 Application of models in real world data

To test the models, in addition to using unseen test data, a further 55 job ads were collected. Fifty job ads were from the Lithuanian "Work Abroad" site, and five ads were collected from industries that were marked as highly present in suspected human trafficking-related advertisements, which were collected through a Google search.

Refer to Table 4.0; the LR model identified two ads as legal job advertisements and two ads as HT-related ads. Similarly, education may represent a higher job requirement in legal job ads, but it is not sufficient to make clear distinctions on how certain words represent respective classes.

Table 4.0

Logistic Regression Performance	Class 1	Class 0
Numbers of ads identified as:	53	2

Table 5.0

Support Vector Machine Performance	Class 1	Class 0
Numbers of ads identified as:	50	5

Commented [JH13]: Table 5

Table 5.0 demonstrates SVM's performance, identifying 5 ads for class 0 and 50 for class 1. By manually observing to distinguish real-world data collected for this testing, SVM's identification of the respective classes is more accurate than the LR model. This suggests that the LR model is overfitting class 1, whereas SVM is able to generalize better.

Chapter 6 Discussion

The project aims to utilize known human trafficking indicators (in the recruitment of victim stage) that are identified in the current research landscape. By manually creating a dataset containing a combination of these indicators, the project examines the machine learning models' capabilities in detecting HT-ads. The findings show that the model performs nearly 'perfectly' on the manually created datasets (train and test). However, in discerning human trafficking-related job ads from additional datasets, SVM outperformed LR models.

This section incorporates theoretical discussion to explain the model performances and sheds light on future directions and improvements for technology-facilitated human trafficking-related research.

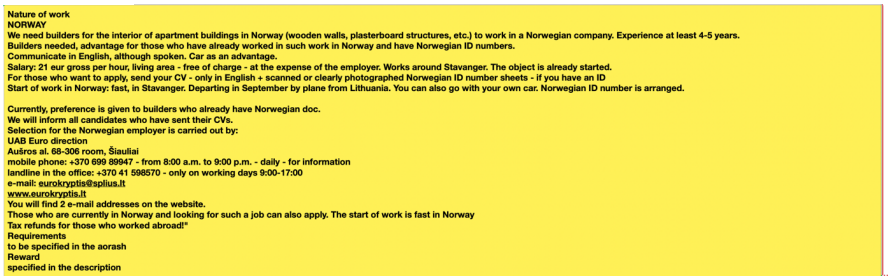
6.1. Methodological challenge

The most difficult challenge in detecting human trafficking-related job advertisements is the availability of datasets confirmed by legal authorities, which corresponds to a broader challenge in HT-related research due to the scarcity of datasets [36]. Specifically, confirming HT-related activity can either rely on victims self-reporting or only be confirmed too late, once the laborer responds to the advertisement and arrives for work [30]. The datasets of potential HT-related job ads used in this research were all collected manually and classified manually at the author's discretion. Therefore, in the final deployment, we cannot ensure with 100% certainty that the classified classes are, in fact, related to HT activities, which highlights a challenge present in this area of research.

6.2 Operationalize Human Trafficking indicator

This dissertation adopts an innovative approach by operationalizing indicators confirmed in various research findings and official reports to develop targeted keywords/phrases vocabularies containing those indicators. The models perform near-perfectly in the controlled datasets. However, when deployed to test real-world data: 50 of the datasets were scraped from darbasuzsienyje.org (workabroad.org), which can contain legal/HT-related job ads, and 4 were scraped from Google searches in high-risk industries: food packaging, truck driving, and waitressing (hospitality). Additionally, 1 ad was directly from Sainsbury Corporation hiring a truck driver. Notably, multiple job ads from workabroad.org claim collaboration with or hiring for Sainsbury food packaging or driving work. The SVM model performs better at discerning ads that belong to class 0; however, its performance in classifying almost all (49 out of 50) job ads from the workabroad.org website into class 1 raises concerns. After manual examination, 3 (out of 50) were determined by the author to belong to class 0, and 2 (out of 50) are open to interpretation. Figure 9.0 provides an example where the author believes it could be classified as class 0 (legal job ads), but the model classified it into class 1.

Figure 9.0


Nature of work
NORWAY
We need builders for the interior of apartment buildings in Norway (wooden walls, plasterboard structures, etc.) to work in a Norwegian company. Experience at least 4-5 years.
Builders needed, advantage for those who have already worked in such work in Norway and have Norwegian ID numbers.
Communicate in English, although spoken. Car as an advantage.
Salary: 21 eur gross per hour, living area - free of charge - at the expense of the employer. Works around Stavanger. The object is already started.
For those who want to apply, send your CV - only in English + scanned or clearly photographed Norwegian ID number sheets - if you have an ID
Start of work in Norway: fast, in Stavanger. Departing in September by plane from Lithuania. You can also go with your own car. Norwegian ID number is arranged.

Currently, preference is given to builders who already have Norwegian doc.
We will inform all candidates who have sent their CVs.
Selection for the Norwegian employer is carried out by:
UAB Euro direction
Aušros al. 68-306 room, Šiauliai
mobile phone: +370 699 89947 - from 8:00 a.m. to 9:00 p.m. - daily - for information
landline in the office: +370 41 598570 - only on working days 9:00-17:00
e-mail: eurokryptis@splius.lt
www.eurokryptis.lt
You will find 2 e-mail addresses on the website.
Those who are currently in Norway and looking for such a job can also apply. The start of work is fast in Norway
Tax refunds for those who worked abroad!°
Requirements
to be specified in the aorash
Reward
specified in the description

Commented [JH14]: Figure 8

The reason the model classifies all job ads from the workabroad.org website can be: 1. Keywords/phrases collected from the website, although screened, still show that almost all job ads on these websites contain at least one indicator related to HT activities. [32] This suggests that the targeted users of this website may be vulnerable to deception by traffickers. The job descriptions can be reasonable for job seekers who wish to work abroad; therefore, accommodations, travel, and relevant documents arranged by employers can be common features in this job market.

To operationalize the indicators in an environment prone to being targeted by traffickers, additional studies, such as interviewing victims specifically deceived by this website into forced labour or gathering experiences from workers abroad originating from Lithuania, would be necessary. Tailoring and operationalizing indicators according to specific data collected from a particular job market in relation to specific demographics and their geographical origin would provide more accurate data and better tailor the training of machine learning models. The keywords and indicators can be operationalized as definitions to red-flag job ads when they present certain indicators, and to set a bar for online job advertisement agencies, requiring them to meet certain requirements and provide employment information to be eligible to post any advertisements.

6.3. Sociological Factors

Technology has become a double-edged sword in the world of combating and enabling technology-facilitated human trafficking activities. Online technologies are expected to provide empirical and actionable information to combat TF-HT activities, protect victims, and prevent HT activities at an early stage (recruitment). [30]
However, the sociological factors that cannot be ignored in the complex issue surrounding HT structures do restrain the usage of technologies in prevention and protection. In this project, the dataset contains HT indicators collected from a Lithuanian job board, and legal job ads were collected from the UK LinkedIn website, where it was verified that they did not contain any indicators. These clean, clear distinctions allow machine learning models to perform well. However, considering the two different socio-economic and cultural/social environments between the Lithuanian job market, advertised towards potential workers relocating, and the UK job market on LinkedIn, the distinction that is not quantified within these two different job markets is reflected in the dataset. Refer to figure 6; the features of class 1 and class 0 are completely different from one another. However, this highlights the

necessity of focusing on detecting indicators within the same job market and socio-economic environment. Human trafficking activities are hidden by nature; therefore, finding subtle indicators within the same environments would provide more valuable actionable insights.

6.4 Technology Facilitated Disinformation strategy and in relations to Human trafficking (recruitment)

Examine the findings in the TF-DI strategy, where human and technological factors further facilitate the spread of disinformation online. These factors and strategies are also presented in the case studies and utilized by traffickers in their DI strategy.

By tackling TF-HT (recruiting victims via the online job market and advertisements) from the perspective of TF-DI, the strategies are seen as cyber-enabled crime, where traditionally human trafficking activities occurred in physical space, but now technology allows them to expand into the online environment.

While there has been a rise in technology-facilitated human trafficking activities, human trafficking arises from a complex set of economic, social, and cultural causes that existed before advancements in technology. [30] Therefore, when researching TF-HT crimes, it is crucial to examine the root causes and then design technology-driven solutions around those complex causes. On the other hand, examining the effects of technology itself can shed new light on how it affects the spread of disinformation in job ads online. For example, the spread of TF-DI results from human factors, such as information glut due to a lack of ground truth and educational awareness, which can lead to false information surrounding human trafficking issues. The enabled environments of online job-seeking platforms could potentially subject vulnerable communities to being targeted by traffickers. The lack of policies and checks and balances around the operation of online platforms poses significant risks to job seekers.

Countermeasures to consider include peer-supplied labels on certain job ads from job seekers to provide further truth and credibility to certain ads. In addition, with an appropriate dataset, machine learning models such as SVM and deep learning models can be used to detect the salient topics and words present in human trafficking ads. Meta-data around HT datasets should be a common practice to allow further discovery of HT-related activities.

Conclusion

The dissertation begins to uncover the technological effect on misinformation and disinformation. By identifying several phenomena that facilitate the rise of misinformation and disinformation in the online space, the research delves deeper through a case study into how human traffickers utilize technology-facilitated disinformation strategies by spreading false job recruitment advertisements online. Difficulties around this research result from the under-reporting of human trafficking issues and the scarcity of real-world data to conduct further research. Despite the challenges, the dissertation aims to operationalize human trafficking-related indicators, identified through conducting thematic literature reviews, and use these indicators to form a specific lexicon vocabulary to feed into the data generation tool (generatedata.com). Then, using the generated datasets to train a Logistic Regression model and a Support Vector Machine (SVM) model, the aim is to test whether the indicators can be operationalized and applied to a machine learning model to contribute to the detection and prevention of the recruitment of vulnerable communities by organized human trafficking organizations.

The findings demonstrate that the SVM shows a degree of discerning capabilities between human trafficking-related ads and legal job ads. However, when the model is deployed on real-world data, it cannot distinguish the ads scraped from a Lithuanian job board for workers seeking jobs abroad. In research by Volodko et al. [32], the authors found at least one indicator of human trafficking present among 430 job posts. This sparks additional considerations, as human trafficking activities are subtle by nature to avoid detection, and workers seeking jobs abroad can be a targeted victim pool for traffickers. Therefore, the dissertation suggests future studies incorporate research designs specific to certain job markets and demographics. If resources permit, it would be beneficial to collect individual experiences and understand the process of getting a job abroad, in addition to creating a demographic profile that is vulnerable to trafficking. This would help in preventing and raising awareness, especially among immigrant worker communities.

Another important discovery in the research is an understanding of the role technology plays in facilitating human trafficking activities. The issues of human trafficking predate the innovation of technology [20]; therefore, researchers should be mindful not to assign the rise of trafficking solely to technology but rather to explore how technology plays a role in the expansion of trafficking activities. Furthermore, it is important to consider whether other socio-economic factors are at play, such as the profitability of human trafficking and the policies of online platforms, like job-seeking websites, which should be stricter and provide sufficient employer information to allow informed decisions by job seekers and to enable law enforcement to trace suspicious activities. The operationalization of the indicators can be used to flag ads if a job advertisement does not meet standards or provide truthful labels on job ads for additional information for job seekers and website users.

Combatting technology-facilitated human trafficking activities requires a multidisciplinary effort, most importantly addressing the stigma around human trafficking challenges, such as underreporting, bias around sexual exploitation, and a lack of awareness of labor trafficking. Additionally, the innovation of technology and its expansion into various aspects of people's lives require stricter policies around the information posted online to keep people safe.

Bibliography:

[1]
T. Tran, Rohit Valecha, and H. Raghav Rao, “Machine and human roles for mitigation of misinformation harms during crises: An activity theory conceptualization and validation,” *International Journal of Information Management*, vol. 70, pp. 102627–102627, Jun. 2023, doi: https://doi.org/10.1016/j.ijinfomgt.2023.102627.
[2]
Y. R. Fung, K.-H. Huang, P. Nakov, and H. Ji, “The Battlefront of Combating Misinformation and Coping with Media Bias,” *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Aug. 2022, doi: https://doi.org/10.1145/3534678.3542615.
[3]
M. Sudhakar and K. P. Kaliyamurthie, “Detection of fake news from social media using support vector machine learning algorithms,” *Measurement: Sensors*, vol. 32, p. 101028, Apr. 2024, doi: https://doi.org/10.1016/j.measen.2024.101028.
[4]
Saumya Pareek and J. Goncalves, “Peer–supplied credibility labels as an online misinformation intervention,” *International journal of human-computer studies*, pp. 103276–103276, Apr. 2024, doi: https://doi.org/10.1016/j.ijhcs.2024.103276.
[5]
UNESCO and J. Posetti, “Balancing act: Countering Digital Disinformation While Respecting Freedom of expression: Broadband Commission Research Report on ‘Freedom of Expression and Addressing Disinformation on the Internet,’” *Unesco.org*, 2020. https://unesdoc.unesco.org/ark:/48223/pf0000379015
[6]
S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” *Science*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018, doi: https://doi.org/10.1126/science.aap9559.
[7]
D. Xu, S. Fan, and Mohan Kankanhalli, “Combating Misinformation in the Era of Generative AI Models,” Oct. 2023, doi: https://doi.org/10.1145/3581783.3612704.
[8]
Indu V and S. M. Thampi, “Misinformation detection in social networks using emotion analysis and user behavior analysis,” *Pattern recognition letters*, vol. 182, pp. 60–66, Jun. 2024, doi: https://doi.org/10.1016/j.patrec.2024.04.007.
[9]
R. Kumari, N. Ashok, T. Ghosal, and A. Ekbal, “Misinformation detection using multitask learning with mutual learning for novelty detection and emotion recognition,” *Information Processing & Management*, vol. 58, no. 5, p. 102631, Sep. 2021, doi: https://doi.org/10.1016/j.ipm.2021.102631.
[10]
G. Kranz, “What Is Metadata and How Does It work?,” *WhatIs.com*, Jul. 2021. https://www.techtarget.com/whatis/definition/metadata
[11]
Jumana Jouhar, A. Pratap, Neharin Tijo, and M. Mony, “Fake News Detection using Python and Machine Learning,” *Procedia computer science*, vol. 233, pp. 763–771, Jan. 2024, doi: https://doi.org/10.1016/j.procs.2024.03.265.
[12]
U. Nations, “Countering Disinformation,” *United Nations*, 2021. https://www.un.org/en/countering-disinformation
[13]

T. Tropina, “Cybercrime and organized crime,” *Freedom from Fear*, vol. 2010, no. 7, pp. 16–17, Jul. 2010, doi: https://doi.org/10.18356/b6f96e06-en.
[14]
R. K. Garrett, “The ‘Echo Chamber’ Distraction: Disinformation Campaigns are the Problem, Not Audience Fragmentation,” *Journal of Applied Research in Memory and Cognition*, vol. 6, no. 4, pp. 370–376, Dec. 2017, doi: https://doi.org/10.1016/j.jarmac.2017.09.011.
[15]
Y. Chai, Y. L. Liu, W. Li, B. Zhu, H. Liu, and Y. Jiang, “An interpretable wide and deep model for online disinformation detection,” 2023.
[16]
A. Ahmad, J. Webb, K. C. Desouza, and J. Boorman, “Strategically-motivated advanced persistent threat: Definition, process, tactics and a disinformation model of counterattack,” *Computers & Security*, vol. 86, pp. 402–418, Sep. 2019, doi: https://doi.org/10.1016/j.cose.2019.07.001.
[17]
D. J. Davis and T. E. Beck, “How social media disrupts institutions: Exploring the intersection of online disinformation, digital materiality and field-level change,” *Information and Organization*, vol. 33, no. 4, p. 100488, Dec. 2023, doi: https://doi.org/10.1016/j.infoandorg.2023.100488.
[18]
A. Petrosyan, “Internet and social media users in the world 2024,” *Statista*, May 22, 2024. https://www.statista.com/statistics/617136/digital-population-worldwide/#:~:text=As%20of%20April%202024%2C%20there
[19]
IBM, “What are AI hallucinations? | IBM,” *www.ibm.com*, 2023. https://www.ibm.com/topics/ai-hallucinations
[20]
D. Barman, Z. Guo, and O. Conlan, “The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination,” *Machine Learning with Applications*, p. 100545, Mar. 2024, doi: https://doi.org/10.1016/j.mlwa.2024.100545.
[21]
B. Anthony, “On-Ramps, Intersections, and Exit Routes: A Roadmap for Systems and Industries to Prevent and Disrupt Human Trafficking SOCIAL MEDIA,” Jul. 2018. Available: https://polarisproject.org/wp-content/uploads/2018/08/A-Roadmap-for-Systems-and-Industries-to-Prevent-and-Disrupt-Human-Trafficking-Social-Media.pdf
[22]
J. Prakash, T. B. Erickson, and H. Stoklosa, “Human trafficking and the growing malady of disinformation,” *Frontiers in Public Health*, vol. 10, Sep. 2022, doi: https://doi.org/10.3389/fpubh.2022.987159.
[23]
“GCHQ | Pioneering a New National Security: The Ethics of Artificial Intelligence,” *www.gchq.gov.uk*. https://www.gchq.gov.uk/artificial-intelligence/index.html
[24]
B. Brewster, T. Ingle, and G. Rankin, “Crawling Open-Source Data for Indicators of Human Trafficking,” *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*, Dec. 2014, doi: https://doi.org/10.1109/ucc.2014.116.
[25]

C. Vajiac *et al.*, “DeltaShield: Information Theory for Human- Trafficking Detection,” *ACM Transactions on Knowledge Discovery From Data*, vol. 17, no. 2, pp. 1–27, Feb. 2023, doi: https://doi.org/10.1145/3563040.
[26]
R. Kapoor, Mayank Kejriwal, and P. Szekely, “Using contexts and constraints for improved geotagging of human trafficking webpages,” *arXiv (Cornell University)*, May 2017, doi: https://doi.org/10.1145/3080546.3080547.
[27]
G. Tyldum and A. Brunovskis, “Describing the Unobserved: Methodological Challenges in Empirical Studies on Human Trafficking,” *International Migration*, vol. 43, no. 1–2, pp. 17–34, Jan. 2005, doi: https://doi.org/10.1111/j.0020-7985.2005.00310.x.
[28]
Department of Homeland Security, “Countering Human Traffcking: Year in Review,” Jan. 2022. Available: https://www.dhs.gov/sites/default/files/2022-02/CCHT%20Annual%20Report.pdf
[29]
“Trafficking in human beings and the internet | Europol,” *Europol*, 2022. https://www.europol.europa.eu/publications-events/publications/trafficking-in-human-beings-and-internet#downloads (accessed Aug. 29, 2024).
[30]
M. Latonero, “Human Trafficking Online: The Role of Social Networking Sites and Online Classifieds,” *SSRN Electronic Journal*, 2011, doi: https://doi.org/10.2139/ssrn.2045851.
[31]
Europol, “Situation Report Trafficking in human beings in the EU,” 2016. Available: https://www.europol.europa.eu/sites/default/files/documents/thb_situational_report_-_europol.pdf
[32]
A. Volodko, E. Cockbain, and B. Kleinberg, “‘Spotting the signs’ of trafficking recruitment online: exploring the characteristics of advertisements targeted at migrant job-seekers,” *Trends in Organized Crime*, vol. 23, no. 1, pp. 7–35, Dec. 2019, doi: https://doi.org/10.1007/s12117-019-09376-5.
[33]
“From Fake Job Ads to Human Trafficking The Horrifying Reality of the Human Trafficking Scam Trade.” Available: https://themekongclub.org/wp-content/uploads/2023/07/From_Fake_Job_Ads_to_Human_Trafficking_The_Horrifying_Reality_of_the_Human_Trafficking_Scam_Trade_2023.pdf
[34]
D. Janušauskienė, “LITHUANIAN MIGRANTS AS VICTIMS OF HUMAN TRAFFICKING FOR FORCED LABOUR AND LABOUR EXPLOITATION ABROAD.” Available: https://humantraffickingsearch.org/wp-content/uploads/2017/06/HEUNI_report_75_15102013.5.pdf
[35]
S. Zhou, J. Peng, and E. Ferrara, “Tracing the Unseen: Uncovering Human Trafficking Patterns in Job Listings,” *SSRN Electronic Journal*, Jan. 2024, doi: https://doi.org/10.2139/ssrn.4883445.
[36]
A. Adepoju, “Data and Research on Human Trafficking,” *Choice Reviews Online*, vol. 43, no. 11, pp. 6–16, Jul. 2006, doi: https://doi.org/10.5860/choice.43-6602.
[37]

B. Andrees and M. N. J. Linden, “Designing Trafficking Research from a Labour Market Perspective: The ILO Experience1,” *International Migration*, vol. 43, no. 1–2, pp. 55–73, Jan. 2005, doi: https://doi.org/10.1111/j.0020-7985.2005.00312.x.
[38]
United Nations, “Protocol to Prevent, Suppress and Punish Trafficking in Persons Especially Women and Children, Supplementing the United Nations Convention against Transnational Organized Crime,” *OHCHR*, Nov. 15, 2000. https://www.ohchr.org/en/instruments-mechanisms/instruments/protocol-prevent-suppress-and-punish-trafficking-persons
[39]
“ONLINE SCAM OPERATIONS AND TRAFFICKING INTO FORCED CRIMINALITY IN SOUTHEAST ASIA: RECOMMENDATIONS FOR A HUMAN RIGHTS RESPONSE.” Accessed: Aug. 30, 2023. [Online]. Available: https://bangkok.ohchr.org/wp-content/uploads/2023/08/ONLINE-SCAM-OPERATIONS-2582023.pdf
[40]
IBM, “What Is Logistic Regression?,” *IBM*, 2024. https://www.ibm.com/topics/logistic-regression
[41]
Towards AI Team, “All About Logistic Regression | Towards AI,” *Towardsai.net*, Mar. 22, 2022. https://towardsai.net/p/machine-learning/all-about-logistic-regression#:~:text=Logistic%20regression%20works%20well%20on (accessed Aug. 29, 2024).
[42]
S. L. Granizo, A. L. Valdivieso Caraguay, L. I. Barona Lopez, and M. Hernandez-alvarez, “Detection of Possible Illicit Messages Using Natural Language Processing and Computer Vision on Twitter and Linked Websites,” *IEEE Access*, vol. 8, pp. 44534–44546, 2020, doi: https://doi.org/10.1109/access.2020.2976530.
[43]
“HUMAN TRAFFICKING INDICATORS,” 2008. Available: https://www.unodc.org/pdf/HT_indicators_E_LOWRES.pdf
[44]
“Contemporary Slavery in the UK. Overview and Key Issues | La Strada Documentation Center about Human Trafficking,” *Lastradainternational.org*, 2024. https://documentation.lastradainternational.org/doc-center/1395/contemporary-slavery-in-the-uk-overview-and-key-issues (accessed Aug. 29, 2024).