Cyber Social Security (CSS): A Lens on Methods for Extraction of Social Sensor Data[⋆]

Valentina Antoniol[1,†], Fabiana Battista[2,†], Paolo Buono[3,†], Danilo Caivano[3,*,†], Gabriella Calvano[4,†], Giuseppe Campesi[1,†], Giuseppe Cascione[1,†], Antonietta Curci[2,†], Marco de Gemmis[3,†], Vincenzo Gattulli[3,†], Roberto La Scala[5,†], Rosa Scardigno[2,†], Annita Larissa Sciacovelli[6,†], Alessandro Senaldi[1,†], Patrizia Sorianello[4,†] and Vincenzo Tamburrano[3,†]

[1]*Dipartimento di Scienze Politiche, University of Bari Aldo Moro*
[2]*Dipartimento Di FOR.PSI.COM, University of Bari Aldo Moro*
[3]*Dipartimento di Informatica, University of Bari Aldo Moro*
[4]*Dipartimento Di Ricerca E Innovazione Umanistica, University of Bari Aldo Moro*
[6]*Dipartimento di Matematica, University of Bari Aldo Moro*
[5]*Dipartimento di Giurisprudenza, University of Bari Aldo Moro*

Abstract

The combination of data coming from social media, smartphones and from urban sensors can actually enable the ability to carry out in-depth analyzes and understand complex phenomena based on human behavior, opening new scenarios for the development of numerous innovative services and applications. By following this research line, the recent paradigm of Social Sensing further emphasized this vision, since it proposed an integrated model in which users themselves are turned into sensors, entities that produce simple rough information which is processed and aggregated in order to generate some valuable human-based findings obtained through the combination and merge of individual-based data. Therefore, considering this scenario, the research work aims to identify and characterize all open information sources that can be interfaced with applications, useful for detecting and interpreting human behavior and the social context, and through language analysis. It also intends to survey and characterize the affective, cognitive, and executive factors that influence/determine human behavior in the use of new technologies. The international geopolitical scenario will also be traced in order to be able to interpret human and social behaviors correctly.

Keywords

Cyber Social Security, CSS, Human Factors, Education, Social Sensor Data

1. Introduction

Social media has become an integral part of modern society, influencing various aspects of our lives, including urban security. It plays a vital role in urban security by enhancing information dissemination, crowd-sourced reporting, community engagement, and awareness. However, it also introduces privacy concerns and the risk of misinformation.

1. *Information Dissemination and Emergency Response*: One of social media's most significant benefits in urban security is its ability to disseminate information during emergencies rapidly. Platforms like Twitter and Facebook have provided real-time updates during natural disasters, terrorist attacks, and other crises. This enables authorities to communicate more effectively with the public and coordinate emergency responses.

Joint Proceedings of IS-EUD 2025: 10th International Symposium on End-User Development, 16-18 June 2025, Munich, Germany.

[*] Corresponding author.

[†] These authors contributed equally.

valentina.antoniol@uniba.it (V. Antoniol); fabiana.battista@uniba.it (F. Battista); paolo.buono@uniba.it (P. Buono); danilo.caivano@uniba.it (D. Caivano); gabriella.calvanoo@uniba.it (G. Calvano); giuseppe.campesi@uniba.it (G. Campesi); giuseppe.cascione@uniba.it (G. Cascione); antonietta.curci@uniba.it (A. Curci); marco.degemmis@uniba.it (M. d. Gemmis); vincenzo.gattulli@uniba.it (V. Gattulli); roberto.lascala@uniba.it (R. L. Scala); rosa.scardigno@uniba.it (R. Scardigno); annitalarissa.sciacovelli@uniba.it (A. L. Sciacovelli); alessandro.senaldi@uniba.it (A. Senaldi); patrizia.sorianello@uniba.it (P. Sorianello); vincenzo.tamburrano@uniba.it (V. Tamburrano)

2. *Crowd-Sourced Reporting*: social media allows citizens to contribute to urban security by reporting suspicious activities, accidents, or emergencies. Mobile applications and hashtags like “See Something, Say Something” encourage people to share information with law enforcement agencies, enhancing situational awareness and overall security in cities.
3. *Community Engagement and Awareness*: Local police departments and city agencies use social media to engage with their communities, share safety tips, and raise awareness about crime prevention.
4. *Surveillance and Privacy Concerns*: While social media aids urban security, it also raises privacy and surveillance concerns. The proliferation of surveillance cameras and the potential for facial recognition technology poses ethical questions about personal privacy.
5. *Misinformation and Panic*: social media can also be a source of misinformation and panic during emergencies. False rumors and fake news can spread rapidly, causing unnecessary alarm and hampering official response efforts.

Social media emerges as a powerful tool for raising awareness about urban violence by allowing survivors to share their stories, build communities, and advocate for change. For instance, platforms like X, played a significant role in the global spread of the #MeToo movement, allowing survivors to share their experiences and raise awareness about the prevalence of sexual harassment and assault. Women’s rights organizations and activists use social media to spread information about resources, support networks, and legal avenues available to survivors. It has facilitated the creation of online communities for survivors to connect and heal. Online campaigns and platforms such as “SafeCity”[1] in India collect crowd-sourced data on incidents of harassment and violence against women, helping to map urban danger zones and advocate for safer cities.

Other forms of violence that can be committed online are cyberbullying and revenge porn, which are often connected. Women who speak out on social media usually face online harassment, threats, and doxxing, which can further victimize them and deter others from reporting incidents. Revenge Porn is the non-consensual sharing of intimate images or videos is another form of online violence against women that is amplified through social media.

In the research, we investigate the role of social media in addressing and amplifying urban violence along the following dimensions:

- Cyber Intimate Partner Violence
- Cyber Gender-based Violence and Stereotypes
- Cyber Hate Speech and Falsehoods
- Urban Mapping & Privacy
- Ethical and Political Risks to interpret Human and Social Behaviors

Figure 1 illustrates a logical architecture and the idea of being able to identify social security units and develop a “Cyber Social Security (CSS)” framework considering the Social Sensor Data [1].

The goal of this research is the proposition of multidisciplinary methods, techniques and tools (IT; psychological, economic, legal, engineering, related to social sciences) capable of operating a Cyber-Social risk management in civil society [2]. To this end, it is necessary to reinterpret the functions of Cyber Security in Cyber Social contexts: **Detection**, **Response**, and **Prevention**.

To achieve this goal, we analyzed different methods for extraction of Social Sensor Data.

2. Security Units in CSS context

Detection: characterize, identify, understand and predict significant cyber-mediated events and changes in human, social, cultural and political behavior as well as the methods for monitoring and protecting ”social” end-points, thus being able to operate with devices (IT and IoT) and diversified information

[1] https://safecity.in/publications/research-papers

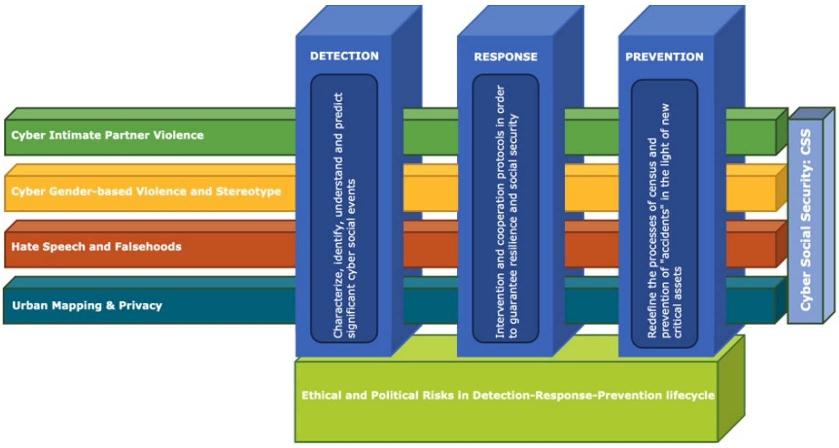
DETECTION
RESPONSE
PREVENTION
Cyber Intimate Partner Violence
Cyber Gender-based Violence and Stereotype
Hate Speech and Falsehoods
Urban Mapping & Privacy
Characterize, identify, understand and predict significant cyber social events
Intervention and cooperation protocols in order to guarantee resilience and social security
Redefine the processes of census and prevention of "accidents" in the light of new critical assets
Cyber Social Security: CSS
Ethical and Political Risks in Detection-Response-Prevention lifecycle

Figure 1: Cyber Social Security Framework

sources (OSINT/CLOSINT) taking into account the national and international legal framework (GDPR, NIS, CyberSecurity Act).

Response: defining intervention and cooperation protocols between the main players in civil society in order to guarantee resilience and social security, including through homeland security technologies and the fight against cyber terrorism and cybercrime. The review of the Detection-Response-Prevention cycle will also clarify the limits within which it is possible to find and manage information while protecting the citizen's right to privacy and the security of civil society.

Prevention: redefine the processes of census and prevention of ”accidents” in the light of new critical assets (individuals, groups, communities, software applications and infrastructures for the public service, etc.), including elements of physical, organizational and applicative security as well as socio-political, economic, psychological and legal context.

3. Methods for detection and interpreting human behavior in CSS and the social context

A range of methods have been developed to detect and interpret human behavior and social context. Villalonga et al. [3] and Baños et al. [4] both propose ontology-based approaches that combine low-level behavior primitives to derive high-level context information. Villalonga's method focuses on activities, locations, and emotions, while Baños extends this to include multimodal context mining.

Groh et al. [5] introduce a method for quantitatively measuring social interactions using infrared tracking, which can be used to identify social situations. Mojarad et al. [6] present a context-aware approach to detecting abnormal human behaviors, using machine learning models to recognize activities, locations, and objects, and an ontology to conceptualize behavior contexts. Onnela et al. [7] highlight the use of sociometers, small sensors that can objectively measure group-level behavior in natural settings. Instead, in [8] is emphasized the use of empirical measurements and mathematical inference to quantitatively analyze individual and social behaviors. Schweizer [9] discusses the use of various research tools, including candidate gene approaches, quantitative genetics, and neuro-endocrine studies, to study the mechanism and function of social behavior. Germain [10] provides an ecological view, considering the influence of various contexts such as society, culture, community, and the physical environment on human behavior.

Therefore, some research works are reviewed to identify the different methods to detect urban violence.

- *Automatic Classification of Sexism in Social Networks: An Empirical Study on Twitter Data* [11]. In response to the escalating spread of hateful and sexist content on social networks, this study introduces a task aimed at understanding and detecting sexism in various forms within online conversations. The study employs traditional and deep learning models for automatic detection by utilizing a newly developed dataset of sexist expressions and attitudes in Spanish on Twitter (MeTwo). Results indicate the prevalence of sexism in diverse forms and the efficacy of deep learning approaches, particularly BERT, in detecting sexist expressions. The study emphasizes the need for nuanced approaches to identify and combat sexism, addressing both explicit hate and subtle stereotypes in online discourse.

- *Domestic violence crisis identification from Facebook posts based on deep learning* [12]. Domestic violence poses a significant threat to public health and human rights. This study addresses the urgent need for quick identification of domestic violence victims through social media. Leveraging deep learning, the study achieves up to 94% accuracy in identifying victims, surpassing traditional machine-learning techniques. The analysis of informative features highlights critical words indicative of crisis situations. The study emphasizes the potential of deep learning in providing timely support to domestic violence victims by automatically identifying crisis situations shared on social me.
- *Modeling stress with social media around incidents of gun violence on college campuses* [13]. Stress is a persistent challenge for college students, exacerbated by violent events on campuses. Leveraging social media as a passive sensor of stress, this study introduces computational techniques to quantify and examine stress responses post-gun violence incidents. A machine learning classifier achieves 82% accuracy in inferring stress expression from Reddit posts. Analyzing social media data around 12 campus gun violence incidents reveals amplified stress levels, characterized by distinctive temporal and linguistic changes. The study highlights the potential of social media analysis in understanding and addressing stress responses following traumatic events on college campuses.
- *Domestic violence and information communication technologies* [14]. The paper addresses the impact of Information Communication Technologies (ICTs) on the experiences of domestic violence survivors, a dimension often overlooked in technological research and design. Through interviews with female survivors residing in a domestic violence shelter, the study reveals the role of mobile phones and social networking sites in post-leaving abuse. Survivors report harassment via mobile phones and experience additional harassment, but also support, through social networking sites. The study underscores the need to consider ICTs in understanding and addressing the challenges faced by domestic violence survivors post-escape.
- *Cyber Aggression and Cyberbullying Identification on Social Networks* [15]. Examining the widespread issue of bullying in the digital realm, this study focuses on cyber aggression and cyberbullying identification on Italian Twitter. Employing Random Forest as the primary classifier, the study processes textual comments to detect aggressive phenomena. The approach achieves notable accuracy and introduces an innovative dataset, the "Aggressive Italian Dataset," providing insights into common patterns in Italian culture related to aggression. The study outlines potential improvements, emphasizing the importance of continuous refinement in identifying and addressing cyberbullying in online social networks.
- *Unveiling Online Sexual Violence Disclosures: A Cross-Platform Analysis Before, During, and After #MeToo* [16]. This study investigates the phenomenon of online disclosure of sexual violence, a relatively recent development. Unlike previous research that predominantly focused on Twitter data during the #MeToo movement, our study spans two years and compares disclosures across various platforms. Using machine learning, we identified 2,927 disclosures for quantitative content analysis. The findings reveal significant differences in timing, information shared, density, co-occurrence, and length across online platforms. Notably, Twitter and the #MeToo movement had the highest number of disclosures, but this study emphasizes the importance of examining online disclosures beyond these contexts. By analyzing Dutch posts, the study aims to reduce heterogeneity and provide a more universal understanding of online disclosures of sexual victimization, addressing cultural and platform-specific factors.
- *Advancing Hate Speech Detection on Social Media: A Transfer Learning Approach* [17], [18], [19]. With the proliferation of social media and the surge in hate speech, detecting offensive content

has become crucial. This study introduces a transfer learning approach for hate speech detection, utilizing pre-trained models for data analysis. Two transfer learning models, Google's Word2vec with LSTM and GloVe with LSTM, are proposed and compared against baseline algorithms. The results demonstrate superior performance in classifying hate, offensive, and neutral speech. The study emphasizes the need for efficient methods to combat hate speech's detrimental societal impacts and highlights the potential of transfer learning in achieving improved detection accuracy across various datasets.

4. Methods for Content Extraction & Annotation

This section aims to describe a framework for the semantic analysis of social streams. It is possible to define a coarse-grained set of requirements which the proposed framework has to implement:

1. Extract textual information from social networks/datasets obtained from social or urban streams.
2. Associate a richer semantics to each piece of content (e.g. the general topic a textual piece of information is about).
3. Associate an opinion (positive, negative, neutral) or a semantic score related to the task being accomplished to each piece of content.
4. Aggregate the information stream in a way that could be exploited to investigate the target phenomenon.

The framework is based on the concept of analysis. Each analysis is run by defining a set of extraction heuristics and some processing steps. In a typical pipeline, a user interacts with the framework by defining the social/urban streams she wants to analyze and the heuristics based on which will then extract content from those streams. Next, the user defines the processes that must be performed on the previously extracted content. The platform's goal is to extract, analyze, aggregate, and organize a very large amount of rough data, in order to produce some valuable analytics for the final user.

The extractor component exploits the APIs to access popular social networks as well as connectors to offline datasets. The resulting database is fed according to specific heuristics (e.g. to extract all the Tweets containing a specific hashtag, all the posts or the Tweets coming from a specific location, and all the posts crawled from specific Facebook pages). After the extraction step, a semantic analysis is performed to associate to each piece of content with the topic it is about.

Given some extraction heuristics, the component connects to social network platforms being investigated to extract content that matches the heuristics and feed the contributions database. The component will implement the bridges towards the platform by exploiting their official APIs. In the case of offline datasets, appropriate "import APIs" will be implemented. For instance, as regards X (formerly Twitter), the content could be extracted by querying the official Streaming APIs, while for Facebook, due to privacy reasons, only the content coming from specific pages or specific groups could be extracted.

In the following, we use the term "document" to refer to a fragment of text in the social stream (e.g. a Facebook or X post, a document in an offline dataset).

However, as extraction heuristics, six different alternatives will be made available in general for a textual social stream:

- Content: extracts all the documents that contain a specific term.
- User: extracts all the documents a specific user posts, given its user name.
- Geo: extracts all the available (geolocalized) documents, given latitude, longitude, and radius.
- Content Geo: extract all the available geolocalized documents that match the terms indicated.
- Page: extract all the posts from a specific page (the main post and the replies).
- Group: extract all the posts from a specific group (the main posts and the replies).

We also plan to work on news and open data. In addition, we exploited the list of relevant RSS feeds for the project to extract news. We downloaded the RSS feeds and ignored those feeds to which it was not possible to connect. Then, the RSS feed processing step is performed: the RSS feed is processed, and for each news item, the GUID (Global unique Identifier) is stored, i.e. the unique identifier of each article. Furthermore, the source of the feed, the date of publication, and the category of the article are stored. For extracting the article from the web page, a Python library for news scraping is available, which parses web pages automatically, extracting the news content. The news will be filtered according to the above-mentioned heuristics. This process continuously feeds the news repository on which we will carry out the analysis according to the purposes of the project.

5. Methods for detecting Cyber Intimate Partner Violence

The proposed domain of Cyber Intimate Partner Violence (C-IPV) intends to investigate the following objectives:

1. studying the phenomenon and the prevalence of C-IPV in Italian population and
2. determining indicators -in terms of personality traits, dispositional differences, and cognitive individuals' characteristics- of this type of violence.

The final aim is to design a predictive model to prevent C-IPV. More precisely, first, it will be conducted a correlational study to assess the prevalence of the main categories of C-IPV (cyber psychological aggression, cyber sexual aggression, and cyber stalking) among Italian population. To better frame the phenomenon, it will be investigated not only the prevalence of victimization but also perpetration rates. It will be also taken into consideration the demographic variable of gender. This is important as so far studies considering this variable have inconclusive results. On one hand, there are studies showing gender differences [20], [21], on the other hand, others did not find differences in males and females [22], [23]. In addition, in this study, individuals' personality and cognitive characteristics ill be tested in order to trace possible correlations, i.e., associations, between the phenomenon of C-IPV and specific individuals' features. The following validated questionnaires will be used to test:

- *Personality and cognitive traits as well as dispositional features*: Dark Triad Dirty Dozen (DTDD), Short Dark Triad (SD3), Assessment of Sadistic Personality (ASP), Trait Alexithymia Scale (TAS-20), Big Five Questionnaire (BFQ), Moral Disengagement Scale (MDS), State-Trait Anger Expression Inventory (STAXI), Ruminative Response Scale (RRS), Anger Rumination Scale (ARS), State-Trait Anxiety Inventory-Y (STAI-Y), Beck Depression Inventory (BDI), Digit Span (DS), Stroop task, Plus Minus task.
- *Cyber intimate partner violence behaviours*: Cyber Dating Violence Inventory (CDVI), Cyber-Dating Abuse Questionnaire (CDAQ) e Cyber Aggression in Relationships Scale (CARS).

In addition, we will also consider sociodemographic variables, such as gender, education, and nationality. Based on the results achieved with the first correlational studies, it will be possible to conceive and deceive further experimental studies to determine human-based behaviors surrounding cyber intimate partner violence.

6. Methods for detecting Hate Speech and Falsehoods

The objective of this domain is to analyse the linguistic characteristics of false statements and hate speech in spoken and written communication. Our focus is on the phonetic and phonological traits of these attitudes, which will be studied using a spectro-acoustic analysis.

Experimental analyses on lying are comparatively sparse; these have primarily focused on the lexical and semantic features of false information within the context of the English language. To date, no acoustic studies have been conducted on the Italian language. In the absence of studies conducted in Italian, it is crucial to carefully select the experimental methodology.

The development of a reliable linguistic approach to lie detection is proving to be an encouraging area of research. When a person is telling a lie, the cognitive load of doing so causes various patterns of speech to emerge. Thus, it is promising to derive a method for predicting truthfulness by analysing speech patterns in comparison to the way a person speaks when telling the truth.

We intend to explore lying in spoken language using controlled tasks such as creating a false story about a personal subject [24], or using a game framework [25]. Given the cognitive demands of lying, a number of prosodic vocal cues appear when someone is lying. According to the results of research conducted on English, liars tend to make more frequent and longer pauses during speech to give themselves more time to construct their lie. Moreover, the act of lying imposes a cognitive load resulting in delayed responses to questions, an increased number of speech errors, and a reduced speech rate. Interestingly, individuals tend to modify their pitch to raise it when lying, with a progressively increasing trend towards the end of each utterance [26].

We aim to examine the phonetic elements of deceptive speech, comparing them with a controlled speech sample. This will involve analysing parameters such as the duration of nuclear syllables, formant frequencies (F0, F1, F2), speech rate, pauses and vowel quality. In the context of prosody, the research aims to investigate the structure of intonation in speech with a focus on the distribution and patterns of pitch accents and boundary tones, as well as pitch range, after extracting the values of f0 min, f0 max, onset and offset.

The second linguistic objective of the project is to examine hate speech. Our particular focus is on the analysis of insults and slurs in written communication: examine the speech acts that lead to the creation of social and emotional tensions between users. Specifically, our objective is to examine how the speaker uses language to intensify insults and increase their illocutionary force. The aim is to investigate the intensification of insults in the context of computer-mediated communication, which is notoriously characterised by mixed and creative means, languages and forms of communication, and which is supplied by paraverbal devices. The research hypothesis posits that individuals who engage in online hate speech are adept at utilizing various linguistic and paralinguistic mechanisms to amplify the impact of their insults, some of which are exclusive to the digital medium. To achieve our goal, a substantial collection of offensive language from various social media platforms will be gathered. This will allow us to analyse the intensification of lexical, semantic, pragmatic, and paralinguistic phenomena in detail.

In terms of spoken language, previous research on impolite speech acts mostly concentrated on pragmatic strategies, neglecting the examination of prosody. The objective of this project is to analyse the prosody of Italian insults and slurs, utilising a spectro-acoustic approach through controlled tasks such as role-plays and the Discourse Completion Task. The aim is to examine the extent to which intensity and pitch range are activated to convey the meaning of invading the hearer's acoustic space. The research hypothesis states that insults exhibit a high intensity and slowed speech rate. The pitch range of the target utterances differs from control speech (either narrowed or widened), as the f0 values are predominantly placed on a high frequency range.

7. Methods for detecting Gender-based Violence and Stereotypes

The opportunities offered by the encounter of methods deriving from social sciences and informatics concern several domains. As a matter of fact, the application of Artificial Intelligence (AI) covers a wide range of domains, including tools to provide critical decisions about who is going to be hired, admitted to college, etc. [27]. As a consequence, it is increasingly important to individuate and eventually mitigate any kind of bias, including gender bias, thus improving fairness in NLP systems. More generally, the increasing presence of online hate groups and web-based hate speech [28], led toward even institutional efforts to contrast those phenomena, such as the Convention on Cybercrime and the Additional Protocol to Regulate Hate Speech Online by the Council of Europe in 2003.

In the research field concerning the extraction of opinions and emotions from text, works on hate groups and in radical forums at the document or sentence level have been usually proposed, as well as lexicon-based semantic content. Analysis of textual data (mostly deriving from the Twitter platform) addressed social, ethnic, sexual or gender minority groups (women, gay and lesbian persons, immigrants, Muslims, Jews and disabled persons) [29]. Despite research in sexism detection represents a growing domain, some of the limitations in the literature about these matters concern the focus on English as the main language and on Twitter as the main platform for data extraction [30]. In addition, sexism embraces a wide range of attitudes and behaviors (such as stereotyping, ideological issues, sexual violence, etc.), and can be expressed in different ways: direct, indirect, descriptive or reported [31], thus implying that misogyny is only one case of sexism [32]. However, studies mostly concentrate on detecting hostile and explicit sexism, while neglecting subtle or implicit expressions of sexism [33], [34].

As a consequence, this research could try to overcome these limits by improving the following research orientations to:

- take into account some specific socio-cultural issues and variables;
- expand the social media sources for data extraction;
- propose a codebook that includes a wide spectrum of sexist attitudes and behaviors, as subtle forms of sexism are most frequent and dangerous for society [35].

The last orientation really represents a critical and core issue, since, to the best of our knowledge, the automatic detection of somewhat implicit content represents a great challenge.

8. International Geopolitical Scenario to interpret the Human and Social Behaviors

The international geopolitical context takes on significant weight in conditioning human and social behavior. In particular, global alliances, political conflicts and economic issues can influence the political, economic, social and cultural conditions that define the environment and context in which people live, transforming their perspectives and priorities. Therefore, although the ”macro level” of geopolitics and the ”micro level” of individual lives clearly refer to two very different scales of analysis, nevertheless they not only can but must be related [36], [37].

A key example of the relationship between geopolitical dimensions and human and social behavior is linked to the issue of fear and security. Ongoing tensions and conflicts, as well as the threat of armed, nuclear or terrorist conflicts, contribute to generating a climate of instability and fear, with physical and psychological repercussions on individuals, prompting them to demand (not necessarily adequate and rational) security and protection measures (Derrida, 1992). In turn, this demand can pave the way for disproportionate measures of social and political control that risk limiting collective and individual freedoms [38].

Not only that, conditions of instability (real or perceived) related to political tensions of local impact, or geopolitical crises of international impact, can also have additional, even very different if not conflicting, consequences. Certainly, they can lead to efforts at international cooperation and solidarity, inducing the development in individuals of a greater awareness of global issues and a sense of responsibility to the international community [39]. At the same time, they can force individuals directly involved in tensions to move (legitimately) to seek protection or better living conditions. Migration phenomena and flows, especially when they are large-scale, can have significant impacts on the communities and territories that welcome (favorably or unfavourably) migrants. Moreover, very often migrants are used as a picklock to stir up fears that are easily exploited politically. It is not uncommon for the very figure of the migrant to be used, by governments and local administrations, as a ”political enemy,” i.e., as a danger that makes the application of control and security measures legitimate, limiting the personal and political freedoms of both migrants and the community at large [40], [41].

It is also important to consider that diplomatic agreements and geopolitical alliances can influence the perception of the “other” and help promote peace on the one hand, but also intensify tensions on the other. In fact, perceived threats from outside can consolidate feelings of belonging and national identity that lead to changes, if not even culture clashes, which can shape public opinions and social values [42]. A key role in these processes is taken on by the media, which amplify geopolitical events, influencing public opinion and contributing (positively or negatively) to the formation of political and cultural positionings. Finally, geopolitical relations, dynamics and tensions have an impact in economic terms, as they influence (including through alliances, trade agreements and economic sanctions) the availability of resources, job opportunities and the cost of living of individuals. They can also lead to more general changes in work patterns, wealth distribution and access to resources.

9. Conclusion

Social media has become a crucial element in enhancing urban security, with its impact spanning various areas such as information dissemination, community engagement, and real-time crisis response. While it significantly contributes to public safety, it also raises concerns around privacy and misinformation.

In the context of urban violence, social media serves as a powerful platform for survivors to share their experiences, build communities, and advocate for change. However, social media also plays a part in online violence, such as cyberbullying and revenge porn, often targeting women. Women who speak out on social media platforms often face online harassment, threats, and doxxing, which not only re-victimizes them but also discourages others from reporting similar incidents. Revenge porn, the non-consensual sharing of intimate images or videos, is another example of online violence amplified through social media.

Therefore, the research focused on understanding the role of social media in urban violence through various dimensions, including cyber intimate partner violence, gender-based violence, cyber hate speech, urban mapping, and privacy concerns. A central aim of the study is the development of a ”Cyber Social Security” framework to address these issues, drawing on multidisciplinary methods, including IT, psychology, economics, law, and social sciences. This framework will focus on the core functions of Cyber Security within social contexts: Detection, Response, and Prevention.

Acknowledgments

This work was partially supported by the following project: SERICS - “Security and Rights in the CyberSpace - SERICS” (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.

Declaration on Generative AI

Either:
The author(s) have not employed any Generative AI tools.

References

[1] V. S. Barletta, M. Calvano, A. Sciacovelli, Cyber social security in multi-domain operations, in: 2024 IEEE International Workshop on Technologies for Defense and Security (TechDefense), 2024, pp. 41–46. doi:10.1109/TechDefense63521.2024.10863352.

[2] M. T. Baldassarre, D. Caivano, B. Fernandez Nieto, D. Gigante, A. Ragone, Fostering human rights in responsible ai: A systematic review for best practices in industry, IEEE Transactions on Artificial Intelligence 6 (2025) 416 – 431. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-86000431264&doi=10.1109%2fTAI.2024.3394389&partnerID=40&md5=8db0230a72f10772e89cc996981dd5bf. doi:10.1109/TAI.2024.3394389.

[3] C. Villalonga, M. A. Razzaq, W. A. Khan, H. Pomares, I. Rojas, S. Lee, O. Baños, Ontology-based high-level context inference for human behavior identification, Sensors (Basel, Switzerland) 16 (2016). URL: https://www.mdpi.com/1424-8220/16/5/645. doi:10.3390/s16050645.

[4] O. Baños, C. Villalonga, J. H. Bang, T. H. Hur, D. U. Kang, S. B. Park, T. Huynh-The, L. Vui, M. B. Amin, M. A. Razzaq, W. A. Khan, C. S. Hong, S. Lee, Human behavior analysis by means of multimodal context mining, Sensors (Basel, Switzerland) 16 (2016).

[5] G. Groh, A. Lehmann, J. Reimers, M. R. Frieß, L. A. Schwarz, Detecting social situations from interaction geometry, in: 2010 IEEE Second International Conference on Social Computing, 2010, pp. 1–8.

[6] R. Mojarad, F. Attal, A. Chibani, Y. Amirat, A context-aware approach to detect abnormal human behaviors, in: ECML/PKDD, 2020.

[7] J. Onnela, B. N. Waber, A. Pentland, S. Schnorf, D. M. Lazer, Using sociometers to quantify social interaction patterns, Scientific Reports 4 (2014).

[8] G. Iacovitti, Quantitative analysis of human behavior, La Clinica terapeutica 161 (2010) 483–484.

[9] D. Schweizer, Social Behaviour Genes Ecology And Evolution, 2016.

[10] C. B. Germain, M. Bloom, Human behavior in the social environment: An ecological view, Columbia University Press, 1999.

[11] F. Rodríguez-Sánchez, J. Carrillo-de Albornoz, L. Plaza, Automatic classification of sexism in social networks: An empirical study on twitter data, IEEE Access 8 (2020) 219563–219576.

[12] S. Subramani, H. Wang, H. Q. Vu, G. Li, Domestic violence crisis identification from facebook posts based on deep learning, IEEE Access 6 (2018) 54075–54085.

[13] K. Saha, M. De Choudhury, Modeling stress with social media around incidents of gun violence on college campuses, Proceedings of the ACM on Human-Computer Interaction 1 (2017) 1–27.

[14] J. P. Dimond, C. Fiesler, A. S. Bruckman, Domestic violence and information communication technologies, Interacting with computers 23 (2011) 413–421.

[15] V. Gattulli, D. Impedovo, G. Pirlo, L. Sarcinella, Cyber aggression and cyberbullying identification on social networks., in: ICPRAM, 2022, pp. 644–651.

[16] M. Gorissen, C. J. W. van den Berg, S. Ruiter, C. C. J. H. Bijleveld, Sharing unwanted sexual experiences online: A cross-platform analysis of disclosures before, during and after the #metoo movement, Computers in Human Behavior 144 (2023) 107724. doi:10.1016/j.chb.2023.107724.

[17] L. Yuan, T. Wang, G. Ferraro, H. Suominen, M. A. Rizoiu, Transfer learning for hate speech detection in social media, arXiv preprint arXiv:1906.03829 (2019).

[18] I. Priyadarshini, S. Sahu, R. Kumar, A transfer learning approach for detecting offensive and hate speech on social media platforms, Multimedia Tools and Applications 82 (2023) 27473–27499. doi:10.1007/s11042-023-14481-3.

[19] R. Ali, U. Farooq, U. Arshad, W. Shahzad, M. O. Beg, Hate speech detection on twitter using transfer learning, Computer Speech & Language 74 (2022) 101365. doi:10.1016/j.csl.2022.101365.

[20] D. C. Bennett, E. L. Guran, M. Ramos, G. Margolin, College students' electronic victimization in friendships and dating relationships: Anticipated distress and associations with risky behaviors, Violence and Victims 26 (2011) 410–429.

[21] P. Leisring, G. Giumetti, Sticks and stones may break my bones, but abusive text messages also hurt: Development and validation of the cyber psychological abuse scale, Partner Abuse 5 (2014) 323–341.

[22] E. Borrajo, M. Gámez-Guadix, E. Calvete, Cyber dating abuse: Prevalence, context, and relationships with offline dating aggression, Psychological Reports: Relationships & Communications 116 (2015) 566–585. (2015a).

[23] J. Temple, H. Choi, M. Brem, C. Wolford-Clevenger, G. Stuart, M. Peskin, J. Elmquist, The temporal association between traditional and cyber dating abuse among adolescents, Journal of Youth Adolescence 45 (2016) 340–349.

[24] M. L. Newman, J. W. Pennebaker, D. S. Berry, J. M. Richard, Lying word: predicting deception from linguistic styles, Personal. Soc. Psychol. Bull. 29 (2003) 665–675.

[25] X. Chen, I. S. Levitan, M. Levine, Acoustic-prosodic and lexical cues to reception and trust: Deciphering how people detect lies, Transactions of the Association for Computational Linguistics 8 (2020) 199–214.

[26] J. Meibauer, The linguistic of lying, Annual Review of Linguistics 4 (2018) 357–375.

[27] J. Doughman, W. Khreich, M. El Gharib, M. Wiss, Z. Berjawi, Gender bias in text: Origin, taxonomy, and implications, in: M. Costa-jussà, H. Gonen, C. Hardmeier, K. Webster (Eds.), Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing, Association for Computational Linguistics, 2021, pp. 34–44. URL: https://aclanthology.org/2021.gebnlp-1.5. doi:10.18653/v1/2021.gebnlp-1.5.

[28] J. Banks, Regulating hate speech online, International Review of Law, Computers & Technology 24 (2010) 233–239. https://ssrn.com/abstract=2129412.

[29] V. Lingiardi, N. Carone, G. Semeraro, C. Musto, M. D'Amico, S. B. and, Mapping twitter hate speech towards social and sexual minorities: a lexicon-based approach to semantic content analysis, Behaviour & Information Technology 39 (2020) 711–721. URL: https://doi.org/10.1080/0144929X.2019.1607903. doi:10.1080/0144929X.2019.1607903. arXiv:https://doi.org/10.1080/0144929X.2019.1607903.

[30] A. Jiang, X. Yang, Y. Liu, A. Zubiaga, Swsr: A chinese dataset and lexicon for online sexism detection, Online Social Networks and Media 27 (2022) 100182. URL: https://www.sciencedirect.com/science/article/pii/S2468696421000604. doi:https://doi.org/10.1016/j.osnem.2021.100182.

[31] M. Hellinger, A. Pauwels (Eds.), Handbook of Language and Communication: Diversity and Change, volume 9 of *Handbooks of Applied Linguistics*, De Gruyter Mouton, 2007. URL: https://www.amazon.it/Handbook-Language-Communication-Diversity-Linguistics-ebook/dp/B01NCZZNIR. doi:10.1515/9783110198539.

[32] K. Manne, Down girl: The logic of misogyny, Oxford University Press, 2017.

[33] Z. Waseem, D. Hovy, Hateful symbols or hateful people? predictive features for hate speech detection on twitter, in: Proceedings of the NAACL Student Research Workshop, Association for Computational Linguistics, San Diego, California, 2016, pp. 88–93. URL: https://aclanthology.org/N16-2013.

[34] S. Frenda, B. Ghanem, M. M. y Gómez, P. Rosso, Online hate speech against women: Automatic identification of misogyny and sexism on twitter, J. Intell. Fuzzy Syst. 36 (2019) 4743–4752. URL: https://api.semanticscholar.org/CorpusID:156056029.

[35] L. Richardson-Self, Woman-hating: On misogyny, sexism, and hate speech, Hypatia 33 (2018) 256–272. doi:10.1111/hypa.12398.

[36] J. Agnew, Geopolitics: Re-visioning World Politics, 2nd ed., Routledge, London, 2003.

[37] L. K. D. Kristofi, The origins and evolution of geopolitics, The Journal of Conflict Resolution 1 (1960) 15–51.

[38] L. Amoore, M. De Goede, Governance, risk and dataveillance in the war on terror, Crime Law and Social Change 43 (2005) 149–173.

[39] M. Kuus, Geopolitics and Expertise: Knowledge and Authority in European Diplomacy, Wiley-Blackwell, Chichester, 2014.

[40] L. Amoore, Biometric borders: Governing mobilities in the war on terror, Political Geography 25 (2006) 336–351.

[41] M. Coleman, A geopolitics of engagement: Neoliberalism, the war on terrorism, and the reconfiguration of us immigration enforcement, Geopolitics 12 (2007) 607–634.

[42] G. Tuathail, Critical Geopolitics: The Politics of Writing Global Space, Routledge, London, 1996.