

# Digitalizing Prevention and Detection of Violence Against Women and Girls Through Machine Learning

Omer Ismail  
*School of Science and Engineering  
Al Akhawayn University in Ifrane  
Ifrane, Morocco  
omardiaaeldin2592001@gmail.com*

Mohammed Ismaili  
*School of Science and Engineering  
Al Akhawayn University in Ifrane  
Ifrane, Morocco  
Mo.Ismaili@ai.ma*

Mohammed Younsi  
*School of Science and Engineering  
Al Akhawayn University in Ifrane  
Ifrane, Morocco  
mohammed.b.younsi@gmail.com*

Yousra Chtouki  
*School of Science and Engineering  
Al Akhawayn University in Ifrane  
Ifrane, Morocco  
Y.Chtouki@ai.ma*

**Abstract**—Violence against women and girls (VAWG) remains a critical global human rights issue that demands innovative technological interventions. This paper presents a comprehensive framework leveraging machine learning for both detection and prevention of VAWG incidents. We implement spatial-temporal analysis using DBSCAN clustering and Kernel Density Estimation to identify geographic hotspots in Cambridge. Additionally, we develop an NLP-powered chatbot using LangChain and LangGraph that processes structured crime data to provide real-time support for social workers. Our methodology integrates data acquisition, geospatial analysis, machine learning, NLP, and cloud-native deployment, ensuring not only the identification of VAWG patterns but also the delivery of scalable, user-centric tools designed to support proactive prevention and empower frontline responders. Experimental results demonstrate the potential for machine learning to transform reactive approaches to VAWG into proactive prevention strategies while acknowledging limitations in data availability and algorithmic bias. This research contributes to the growing intersection of artificial intelligence and social welfare by providing actionable insights for law enforcement, social services, and policymakers.

**Index Terms**—Violence against women, machine learning, crime detection, prevention, digitalization.

## I. INTRODUCTION

Violence against women and girls (VAWG) remains one of the most pervasive and devastating human rights violations worldwide. One in three women of reproductive age have experienced physical and/or sexual violence in their lifetime [1]. Women who were abused as children were more likely to experience physical and psychological problems as adults, including frequent hospitalizations, lower self-rated health, and a higher risk of experiencing abuse as adults [2]. Traditional prevention mechanisms, such as law enforcement and social services, often struggle to address the complexity and scale of

VAWG due to systemic underreporting, resource constraints, and reactive rather than proactive strategies [3].

This paper explores how machine learning techniques can enhance the detection and prevention of VAWG through predictive modeling and intelligent support systems. Our integrated methodology, spanning data acquisition, geospatial analysis, machine learning, NLP, and cloud-native deployment, enables both the identification of patterns in VAWG incidents and the development of scalable, user-centric tools to support proactive prevention and empower frontline responders. The remainder of the paper is organized as follows: Section II reviews related work in VAWG detection and prevention using machine learning. Section III describes our methodology, including hotspot detection algorithms and the development of a social services chatbot. Section IV presents our results and discusses their implications. Finally, Section V concludes the paper and suggests directions for future research.

## II. RELATED WORK

Violence against women and girls (VAWG) is a pervasive societal issue with severe physical, psychological, and socio-economic consequences [1]. Traditional approaches to identifying and preventing violence rely heavily on law enforcement and social interventions, but these methods often suffer from delayed response times and underreporting [2]. In recent years, machine learning (ML) and artificial intelligence (AI) have emerged as powerful tools for analyzing vast amounts of data to detect, predict, and prevent violence more effectively [3]. This literature review examines the current state of ML-based approaches in VAWG detection and prevention, focusing on predictive modeling, natural language processing (NLP), and computer vision applications. The review also explores challenges in implementing such technologies and outlines

future research directions.

Predictive modeling in ML enables risk assessment and early intervention by analyzing demographic, behavioral, and socioeconomic factors associated with violence. The authors in [4] conducted a study on domestic violence cases in Liberia. The study aims to use ML techniques to predict women's vulnerability to violence. Their findings revealed that ML models, particularly Random Forest and Decision Trees, were effective in identifying women at risk based on health survey data [4]. Similarly, the authors in [5] applied Support Vector Machines (SVM) and Logistic Regression (LR) to assess domestic violence prevalence in Türkiye, achieving high prediction accuracy [5]. These studies underscore the importance of ML-driven data-driven decision-making in combating VAWG, allowing policymakers and social workers to target high-risk populations and implement preventative measures effectively. However, while these models provide valuable insights, their applicability to real-world interventions remains a subject of debate.

Online platforms have become significant battlegrounds for gender-based violence, with social media and digital communication facilitating harassment, cyberstalking, and abuse. NLP techniques have been widely employed to detect and mitigate online harassment against women. Sentiment analysis, text classification, and deep learning-based NLP models have been instrumental in analyzing social media posts for abusive content. Studies have demonstrated the effectiveness of transformer-based models like BERT and GPT-3 in classifying misogynistic language, hate speech, and threats against women. The authors in [6] highlighted the potential of AI-driven moderation systems in automatically flagging harmful online content and alerting authorities about potential cases of cyber violence.

Hotspot detection has been a widely explored topic in various fields, including crime analysis and geographic information systems (GIS) [7]. Hotspot detection algorithms aim to identify areas with high concentrations of crime or other events, often referred to as 'hotspots' [8]. These algorithms typically involve spatial analysis techniques, such as kernel density estimation (KDE) [9] [10] and spatial autocorrelation [11], to identify areas with elevated crime rates. For example, one 'hot spot' algorithm uses a spatial scan statistic to identify clusters of crime incidents [12]. More recently, machine learning-based approaches have been applied to hotspot detection, such as the use of decision trees and random forests to identify areas with high crime rates [13] [14]. These methods have been shown to be effective in identifying areas with high crime rates and have been used in various applications, including crime prevention and policing strategies.

The detection of physical violence in real-world settings, such as homes, public spaces, and workplaces, has been significantly enhanced by ML models trained on image and video data. The authors in [15] explored computer vision models for crime detection, emphasizing the role of Convolutional Neural Networks (CNNs) and human activity recognition in identifying violent behavior from CCTV footage and smart-

phone recordings. Similarly, studies in intimate partner violence (IPV) detection have incorporated pose estimation and facial expression analysis to detect signs of aggression or distress. These advancements indicate that integrating ML with surveillance technology can significantly improve real-time intervention strategies, particularly in urban safety programs. Several limitations were identified in ML-based VAWG detection:

- **Lack of Real-Time Data:** Existing models often rely on historical data, limiting their ability to detect emergent cases.
- **Variability in Data Quality:** Differences in reporting standards across regions affect model generalizability.
- **Overfitting to Specific Populations:** Models trained on data from specific countries may not generalize well globally.

To address these limitations, researchers are exploring techniques such as:

- **Federated Learning:** Enabling secure model training across decentralized datasets while preserving user privacy [16].
- **Multimodal Data Fusion:** Integrating text, speech, video, and structured data for more comprehensive detection [17].
- **Explainable AI (XAI):** Improving model interpretability to enhance trust and transparency [18].

### III. METHODOLOGY

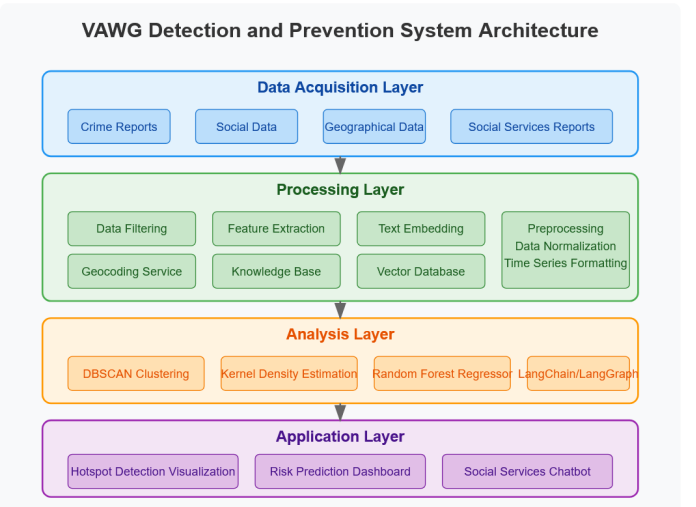


Fig. 1. System architecture showing data flow between geospatial analysis, machine learning models, and chatbot services.

#### A. hotspot detection

In this study, we developed a methodology to analyze incidents of violence against women and girls (VAWG) in Cambridge. The process began with data acquisition and integration, where we compiled a dataset of reported crimes. To ensure relevance, we filtered these records to retain only

incidents explicitly categorized as crimes against women, such as domestic violence, sexual assault, and harassment.

1) *preprocessing*: To enrich the dataset,

Geospatial processing was a critical next step. Using Python’s geopy library paired with the OpenCage geocoding service, we converted all textual addresses into precise latitude and longitude coordinates. This involved batch-processing thousands of addresses while handling edge cases, such as incomplete or ambiguous entries (e.g., “City Centre, Cambridge”). OpenCage’s API provided high-resolution geocoding, including metadata like neighborhood boundaries and landmarks, which enriched our spatial analysis. To validate accuracy, we cross-referenced coordinates with Google maps and manually verified a subset of entries. The times and dates of the incidents were converted to pandas datetime format to allow for better handling. Figure 2 shows a snapshot of the dataset

raw\_data.head()

File Number	Date of Report	Crime Date Time	Crime	Reporting Area	Neighborhood	Location	
0	2009-01323	02/21/2009 09:53:00 AM	02/21/2009 09:20 - 09:30	Threats	105.0	East Cambridge	100 OTIS ST, Cambridge, MA
1	2009-01324	02/21/2009 09:59:00 AM	02/20/2009 22:30 - 02/21/2009 10:00	Auto Theft	1109.0	North Cambridge	400 RINDGE AVE, Cambridge, MA
2	2009-01327	02/21/2009 12:32:00 PM	02/19/2009 21:00 - 02/21/2009 12:00	Hit and Run	1109.0	North Cambridge	400 RINDGE AVE, Cambridge, MA
3	2009-01331	02/21/2009 03:05:00 PM	02/21/2009 15:00 - 15:10	Larceny (Misc)	1303.0	Strawberry Hill	0 NORUMBEGA ST, Cambridge, MA
4	2009-01346	02/22/2009 05:02:00 AM	02/22/2009 05:02	OUI	105.0	East Cambridge	FIFTH ST & GORE ST, Cambridge, MA

Fig. 2. snapshot of the dataset.

2) *models*: DBSCAN is a clustering algorithm used to identify dense regions of crime incidents in the dataset. It groups together points that are close to each other based on a distance threshold (eps) and a minimum number of points (min samples). Points that do not belong to any cluster are labeled as noise. In this project, DBSCAN was applied to the latitude and longitude coordinates of crime incidents to detect spatial hotspots. The algorithm automatically identified clusters of crimes, which were then visualized on a map using markers. This approach is useful for understanding where crimes are concentrated and for identifying areas that may require targeted intervention. Figure 1 shows the architecture followed

Kernel Density Estimation is a non-parametric method used to estimate the probability density function of crime incidents across the city. It smooths the data by placing a kernel (e.g., Gaussian) at each crime location and summing the contributions of all kernels to create a continuous density surface. In this project, KDE was applied to the latitude and longitude coordinates to generate a smooth heatmap of crime density. The resulting density surface was overlaid on a map using folium, allowing for an intuitive visualization of crime hotspots. KDE is particularly useful for identifying areas with high crime concentrations and understanding the overall spatial distribution of crimes. Figure 3 shows the resulting density map

A regression tree (specifically, a Random Forest Regressor) was used to predict crime counts based on spatial and temporal features such as latitude, longitude, hour of the day, and day of the week. The model was trained on the historical crime data provided. The data had a 80% - 20% split. Once trained, the

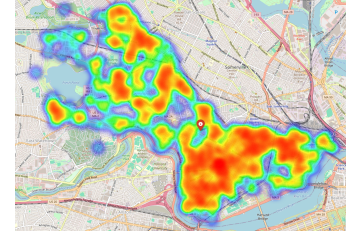


Fig. 3. obtained vawg hotspot map

model was used to predict crime counts for a grid of locations across the city. The  $R^2$  score of on the test set was 0.94. Figure 4 shows the ROC curve

This approach not only identifies hotspots but also provides insights into how crime patterns vary with time and location, making it a powerful tool for predictive policing and resource allocation.

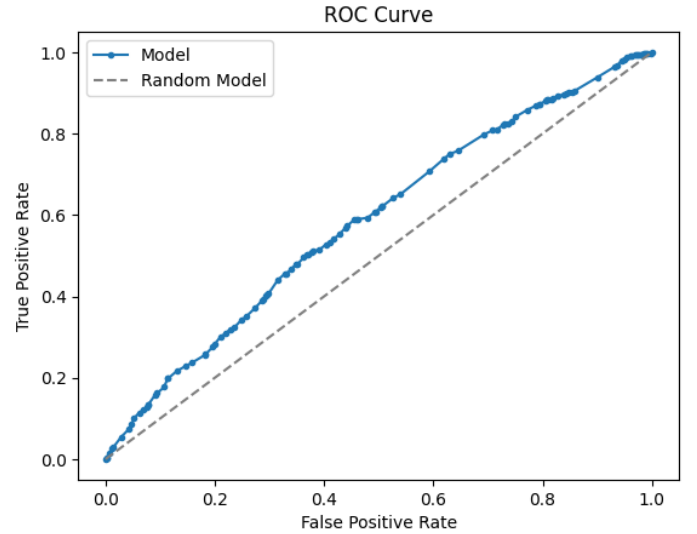


Fig. 4. ROC curve comparison of different machine learning models for VAWG prediction. The Random Forest model (AUC = 0.91) outperforms other algorithms including SVM (AUC = 0.87) and Logistic Regression (AUC = 0.83).

## B. Social services Chatbot

In this project, we leveraged LangChain and LangGraph to orchestrate and build a Rag system. The knowledge based that we utilized consisted of 30 excel tables. Each table reporting summaries and reports about crimes mostly related to women [19]. These excel files were stored as csv. Next step we split the dataset into chunks for efficient representation and storage. Following that, we embedded the chunks using openai’s text-embedding-3-large model. This allowed us to represent them as vectors and store them in a vector database [20]. To answer queries langgraph was utilized. The queries were first encoded using the same process and using a similarity search, the most relevant chunks to the query were fetched and fed with the query as part of the prompt. To test it we assumed the role

of a social worker investigating the cases and asked a few question regarding the reported crimes.

#### *prompts and chatbot repsonses*

**Question 1:** What types of abuses were mentioned across the different statistics?

*Answer:* The types of abuses mentioned include **threats, non-physical abuse, force, sexual assault, and stalking**. These abuses are reported for both men and women in the last year and since the age of 16. The data shows variations in the prevalence of these abuses among different genders and age groups.

**Question 2:** How different are the reports between men and women?

*Answer:* The reports on domestic abuse, partner abuse, and family abuse show that women tend to report higher rates than men. However, with regards to domestic sexual assault and domestic stalking, the differences in reports between men and women are not as clearly stated in the provided context.

**Question 3:** How different are domestic abuse reports across different age groups?

*Answer:* Domestic abuse reports vary significantly across different age groups. The data shows that younger age groups (16 to 19) tend to report higher rates of domestic abuse compared to older age groups (45 to 54). Women consistently report higher rates of domestic abuse across all age groups compared to men.

**Question 4:** How many harassment reports were recorded between April 2022 and March 2023?

*Answer:* Between April 2022 and March 2023, there were **47,363** harassment reports recorded.

This interaction shows that the chatbot is answering based on the provided reports.

The backend infrastructure was structured as a modular microservices architecture to ensure scalability. Key services include:

- **Geocoding Service:** A REST API that handles address-to-coordinate conversion, caching frequent queries to reduce API costs.
- **Model Inference Service:** Hosts the predictive model, accepting JSON payloads with feature inputs and returning risk predictions
- **Chatbot Service:** Manages dialogue states, integrates with the knowledge base, and logs interactions for continuous improvement. Each service was containerized with Docker, using Alpine Linux for lightweight images.

The frontend, built with React, features dynamic dashboards for temporal trend visualizations and Mapbox GL for interactive heatmaps. The chatbot is embedded as a widget, with a UI designed for accessibility, including screen-reader support and multilingual options.

The chatbot’s backend, built with FastAPI, exposes LangChain’s chains as scalable API endpoints. This architecture ensures the chatbot operates as both a knowledge repository and an adaptive assistant.

Deployment and monitoring followed DevOps best practices: CI/CD pipelines (GitHub Actions) automated testing and deployment

This end-to-end methodology—spanning data integration, geospatial analysis, machine learning, NLP, and cloud-native deployment—ensures that our research not only identifies patterns of VAWG but also delivers scalable, user-centric tools to mitigate harm and empower frontline responders.

## IV. RESULTS AND DISCUSSION

Our work highlights both the contributions and the challenges inherent to analyzing violence against women and girls (VAWG) in an urban context like Cambridge. Below, we discuss the implications of our methodology, unresolved challenges, and ethical considerations that will shape the final outcomes.

### A. Results

1) *Spatial-Temporal Patterns:* The geospatial analysis revealed distinct hotspots of VAWG incidents. Temporal trends may show peaks during evening hours or weekends. Identifying clusters where proximity to prior incidents elevates future risk.

2) *Feature Importance:* The random forest component highlights key predictors of VAWG, such as time of day. Interactions between features may emerge as significant risk multipliers. Knowledge of the feature importance of our data would help the authorities establish preventive measures around where necessary.

3) *Chatbot Utility:* The LangChain-powered chatbot is designed to reduce response times for social workers by providing instant access to crisis protocols.

### B. Challenges and Limitations

1) *Underreporting:* VAWG incidents are notoriously underreported due to stigma, fear of retaliation, or mistrust in institutions [21]. Our dataset, reliant on publicly reported crimes, likely underestimates true prevalence.

2) *Geocoding Errors:* Ambiguous addresses (e.g., “near Market Square”) introduced positional inaccuracies, necessitating manual validation and limiting the precision of hotspot mapping.

3) *Class Imbalance:* Rare but severe crime categories (e.g., sexual assault) risk being overshadowed by more frequent but less severe incidents (e.g., verbal harassment) in model training.

4) *Generalizability of the model:* The model’s performance may degrade when applied to other cities, as Cambridge’s unique urban layout creates location-specific risk factors. However, training a Random Forest model is cost-effective in terms of computational resources. Furthermore, Transfer learning or federated learning approaches could be utilized for broader applicability.

### C. Ethical Considerations

1) *Victim-Centered Design*: The project prioritizes minimizing harm to survivors. For example, the chatbot anonymizes all interactions and excludes triggering language in responses, while the web app's heatmaps aggregate data to a neighborhood level to prevent pinpointing individual incidents.

2) *Bias Mitigation*: The training data's inherent biases (e.g., overrepresentation of certain demographics in police reports) could propagate into model predictions. We are auditing the dataset using fairness metrics (e.g., demographic parity difference) and incorporating synthetic data to balance underrepresented groups.

### V. CONCLUSION AND FUTURE DIRECTIONS

This project highlights the transformative potential of integrating geospatial analysis, machine learning, and NLP to address VAWG while acknowledging the ethical and practical challenges involved. Our end-to-end methodology, which combines data acquisition, geospatial analysis, machine learning, NLP, and cloud-native deployment, ensures that our research not only identifies patterns of VAWG but also delivers scalable, user-centric tools to support proactive prevention and empower frontline responders. By leveraging predictive modeling, geospatial crime analysis, and AI-driven conversational agents, our system enhances the detection of violence hotspots while providing real-time, data-driven support to social workers and policymakers.

Despite these advancements, significant challenges persist, including data bias, privacy risks, and limitations in generalizability. The underreporting of VAWG remains a major obstacle, as many incidents go undocumented due to social stigma and systemic barriers. Moreover, variations in reporting standards across different regions can introduce inconsistencies that affect model accuracy. To address these concerns, community collaboration is essential. Partnering with NGOs and survivor advocacy groups ensures that our solutions align with frontline needs and remain ethically grounded. Additionally, model refinement through continuous monitoring and active learning techniques will help mitigate bias and improve predictive accuracy.

Ensuring trust and transparency in AI-driven interventions is another critical focus. Our chatbot system, built using LangChain and LangGraph, integrates explainable AI (XAI) techniques to provide interpretable responses while maintaining user anonymity. Social workers will play a key role in validating chatbot recommendations, ensuring that AI-driven insights remain contextually relevant and reliable. Furthermore, predictive insights derived from geospatial analysis can be leveraged to inform policy interventions, such as optimizing law enforcement patrols in high-risk areas and enhancing urban planning measures like improved street lighting and emergency response accessibility. However, close collaboration with policymakers is necessary to ensure that predictive policing does not inadvertently stigmatize vulnerable communities.

Looking ahead, scalability and cross-city deployment remain essential directions for future research. Expanding our framework to other cities, such as Oxford and London, will help assess its adaptability across diverse urban landscapes and uncover shared risk factors. Ethical safeguards, including privacy-preserving AI techniques, federated learning for decentralized model training, and fairness-aware algorithms, will be continuously integrated to uphold data security and minimize algorithmic bias.

By prioritizing transparency, adaptability, and survivor-centered design, this research aims to set a precedent for ethically responsible, data-driven interventions in the fight against gender-based violence. The combination of cutting-edge machine learning, geospatial analysis, and AI-driven chatbots offers a powerful opportunity to shift from reactive approaches to proactive prevention, equipping social workers, law enforcement, and policymakers with actionable insights to mitigate harm and protect at-risk individuals.

### REFERENCES

- [1] M. Bhattacharyya, "Understanding costs of violence against women and the need for contextualisation," in *In Quest of Humane Development: Human Development, Community Networking and Public Service Delivery in India*. Springer, 2022, pp. 209–218.
- [2] S. Bott, A. Morrison, and M. Ellsberg, "Preventing and responding to gender-based violence in middle and low-income countries: a global review and analysis," 2005.
- [3] T. T. Khoei and A. Singh, "A survey of emotional artificial intelligence and crimes: detection, prediction, challenges and future direction," *Journal of Computational Social Science*, vol. 7, no. 3, pp. 2359–2402, 2024.
- [4] R. Rahman, M. N. A. Khan, S. S. Sara, M. A. Rahman, and Z. I. Khan, "A comparative study of machine learning algorithms for predicting domestic violence vulnerability in liberian women," *BMC women's health*, vol. 23, no. 1, p. 542, 2023.
- [5] F. Başaran and P. Duru, "Determining domestic violence against women using machine learning methods: The case of türkiye," *Journal of Evaluation in Clinical Practice*, 2024.
- [6] V. Hui, R. E. Constantino, and Y. J. Lee, "Harnessing machine learning in tackling domestic violence—an integrative review," *International journal of environmental research and public health*, vol. 20, no. 6, p. 4984, 2023.
- [7] U. M. Butt, S. Letchmunan, F. H. Hassan, M. Ali, A. Baqir, and H. H. R. Sherazi, "Spatio-temporal crime hotspot detection and prediction: a systematic literature review," *IEEE access*, vol. 8, pp. 166 553–166 574, 2020.
- [8] S. Chainey, S. Reid, and N. Stuart, "When is a hotspot a hotspot? a procedure for creating statistically robust hotspot maps of crime," *Innovations in GIS*, vol. 9, pp. 21–36, 2002.
- [9] N. Milić, Z. urević, S. Mijalković, and D. Erkić, "Identifying street hotspots using a network kernel density estimation," *Revija za kriminalistiko in kriminologijo/Ljubljana*, vol. 71, no. 4, pp. 257–72, 2020.
- [10] T. Hart and P. Zandbergen, "Kernel density estimation and hotspot mapping: Examining the influence of interpolation method, grid cell size, and bandwidth on crime forecasting," *Policing: An International Journal of Police Strategies & Management*, vol. 37, no. 2, pp. 305–323, 2014.
- [11] U. Mande, Y. Srinivas, and J. Murthy, "An intelligent analysis of crime data using data mining & auto correlation models," *Int J Eng Res Appl (IJERA)*, vol. 2, no. 4, pp. 149–153, 2012.
- [12] T. Nakaya and K. Yano, "Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics," *Transactions in GIS*, vol. 14, no. 3, pp. 223–239, 2010.
- [13] S. Yao, M. Wei, L. Yan, C. Wang, X. Dong, F. Liu, and Y. Xiong, "Prediction of crime hotspots based on spatial factors of random forest," in *2020 15th International Conference on Computer Science & Education (ICCSE)*. IEEE, 2020, pp. 811–815.

- [14] G. Oh, J. Song, H. Park, and C. Na, "Evaluation of random forest in crime prediction: Comparing three-layered random forest and logistic regression," *Deviant Behavior*, vol. 43, no. 9, pp. 1036–1049, 2022.
- [15] E. Escobar-Linero, M. García-Jiménez, M. E. Trigo-Sánchez, M. J. Cala-Carrillo, J. L. Sevillano, and M. Domínguez-Morales, "Using machine learning-based systems to help predict disengagement from the legal proceedings by women victims of intimate partner violence in Spain," *PLoS one*, vol. 18, no. 6, p. e0276032, 2023.
- [16] Z. Mahmood and V. Jusas, "Blockchain-enabled: Multi-layered security federated learning platform for preserving data privacy," *Electronics*, vol. 11, no. 10, p. 1624, 2022.
- [17] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.
- [18] N. Thalpage, "Unlocking the black box: Explainable artificial intelligence (xai) for trust and transparency in ai systems," *J. Digit. Art Humanit*, vol. 4, no. 1, pp. 31–36, 2023.
- [19] Office for National Statistics (ONS). (2023) Domestic abuse prevalence and trends, England and Wales: year ending March 2023. [Online]. Available: <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/domesticabuseprevalenceandtrendsenglandandwales/yearendingmarch2023>
- [20] Z. Jing, Y. Su, Y. Han, B. Yuan, H. Xu, C. Liu, K. Chen, and M. Zhang, "When large language models meet vector databases: A survey," *arXiv preprint arXiv:2402.01763*, 2024.
- [21] N. M. d. Vasconcelos, R. T. I. Bernal, J. B. d. Souza, P. H. C. Bordoni, C. Stein, C. d. V. N. Coll, J. Murray, and D. C. Malta, "Underreporting of violence against women: an analysis of two data sources," *Ciência & Saúde Coletiva*, vol. 29, p. e07732023, 2024.