Using AI to reduce Online Gender Based Violence

By
Nelly Nyadzua
Artificial Intelligence and Cyber Security Specialist

Twitter: @NellyNyadzua
LinkedIn: Nelly Nyadzua
Phone: **+254751086176**

ONLINE GENDER BASED VIOLENCE

- Online Gender Based Violence is targeted harassment and prejudice through technology against people, disproportionately women, based on their gender.
- According to Association for Progressive Communications (APC), 73% of adult online users have seen someone being harassed online, 40% have personally experienced a form of online harassment

Forms of online harassment

- **Infringement of Privacy** through cracking personal accounts, doxing, sharing sexualized clips, revenge porn.
- **Surveillance and Monitoring** as Stalking, Spying Key loggers, GPS trackers
- **Harassment** through Cyberbullying, sexist comments, hate speech, Direct threats of physical and sexual violence
- **Damaging reputation and credibility** as defamation, identity theft.
- **Direct threat/ violence** as Sexualized blackmail, advocating femicide, impersonating resulting to physical attack

OGBV Intent and Target

Women and girls are frequently victims of gender-based violence in virtual environments, and online abuse has become a major social issue. High-profile women in a variety of professions are frequently targets of online violence.

Intentions of OGBV

- Silence an opinion
- Drive someone out of digital space
- Exclude people from debates/ decision making processes
- Restrict socio-economic, cultural and political participation
- Lower financial income of the target
- discredit/ stain the reputation of target

Human rights principles that provide protection against GBV include

- Privacy rights
- Consumer protection rights
- Laws against defamation
- Laws against hate speech

Harm of OGBV

Online GBV intrudes women's right to self-determination and bodily integrity, impacts on their capacity to move freely, without fear of surveillance, and denies them the opportunity to craft their own identities online, and to form and engage in socially and politically meaningful interactions.

- **Psychological harm** through which victims/survivors experience depression, anxiety and fear
- **Social isolation** through which victims/survivors withdrew from public life, including with family and friends.
- **Economic loss** through which victims/survivors became unemployed and lost income.
- **Limited mobility** through which victims/survivors lost the ability to move around freely and participate in online and/or offline spaces.
- **Self-censorship** for fear of further victimization and due to loss of trust in the safety of using digital technologies

Case Study: Kenya victims on Twitter

Kenyans on Twitter (KOT)– A group of accounts and bots owned by some Kenyans who rally the regular users/ accounts on twitter to talk about a certain topic and give it much concentration to make it trend.

<u>Political Victims of KOT</u>

- Former President Uhuru Kenyatta- Closed his account , over 25 parody accounts created
- Former Cabinet Secretary for the Ministry of interior and Coordination of National Government Fred Matiang'i closed account and 10 parody accounts came up
- Female Politician , Daughter of Former Prime Minister- Winnie Odinga
- Lady Justice Martha Koome
- Nairobi County Woman Representative Esther Passaris
- Over 30 female politicians – 70%(21) closed accounts, 20% (6) paused social media overall, 10% (3) kept the accounts

Online Regulations

- Current Regulations in place are:
- Computer misuse and Cybercrimes Act , Kenya
- International human rights law
- International Covenant on Civil and Political Rights (ICCPR)
- Twitter Rules and Policies
- Other Social media platforms rules and policies eg. Meta, IG, Tiktok

Common Barriers

- Clash of Rights
- **Dataset Language barriers-** datasets in english yet offence happens in Swahili, mother-tongue and slang words that AI doesn't recognize
- **Legal Liability** – Terms of service binging company's legal obligation in Country of residence.
- Anonymity, VPN and encryption to protect perpetrators

How AI can help track OGBV

- **Control data in volumes-** AI can be used to moderate uploaded content and monitor user interactions on social sites, which would be impossible to do manually due to the volume.
- **Deletion of Content and Accounts-** Algorithms are programmed to sift through massive amounts of content and delete both posts and users when the content is harmful and does not adhere to platform standards.
- **Sentiment Analysis-** Constantly evolving and learning algorithms, gaining ability of recognizing duplicate posts, comprehending the context of scenes in videos, and even recognizing tones such as anger or sarcasm. If a post is unable to be identified, it will be marked for human review.
- **Content Moderation-** Reviewing majority of online activity protects human moderators from disturbing content that could otherwise lead to mental health issues.
- **Protect Vulnerable targets-** Natural Language Processing (NLP) tools to monitor interactions between users and identify inappropriate messages being sent amongst underage and vulnerable users.

How AI can help track OGBV

- **Censor attackers-** Identify malicious users and harmful content generated by a minority of users and use AI techniques to prioritize their content for review.
- **Pattern Recognition-** Machine learning enables these systems to find patterns in behaviors and conversations invisible to humans and can suggest new categories for further investigation.
- **Verification Automation-** Use AI to verify information and the validate of a post's authenticity to eliminate the spread of misinformation and misleading content.
- AI can be used to **proactively educate** users about responsible online behavior through real-time alerts and blockers.
- Prevent sharing of personal information or inappropriate messages by intervening in real-time

Challenge to you!

- Further papers have been written to address the issue of OGBV such as the Deep learning neural network for detection of Gender-Based Violence in Twitter messages in Mexico - https://link.springer.com/chapter/10.1007/978-3-030-89691-1_3
- Above are highlights on how AI can be used to address and control the occurrences of OGBV but much more action is needed.
- A multifaceted approach to using AI to address, reduce and solve online violence against women is needed through:

- **Preventive** (including education and technical features, for example)
- **Reactive** (swiftly take down unlawful content, and investigate)
- **Redress** for victims.

- I challenge you to seek practical and sustainable AI solutions to address this monster that is crawling in the shadows!


Thank You