Countering online hate speech: A qualitative study of digital literacy training programs

Yasmin Osama

ملخص الدراسة

تستكشف هذه الدراسة الكيفية فعالية البرامج التدريبية للتربية الإعلامية الرقمية في مواجهة خطاب الكراهية عبر المنصات الرقمية، من خلال مقابلات متعمقة شبه منظمة مع 10 من صانعي المحتوى الذين شاركوا في برنامج تدريبي وجاهي في مركز تدريب راديو هولندا RNTC في الفترة من سبتمبر– أكتوبر 2022، بعنوان "الإعلام الرقمي في مواجهة المعلومات المضللة وخطاب الكراهية". وتشير النتائج إلى أن المشاركين وجدوا تطورا ملحوظا في كفاءتهم في تحديد خطاب الكراهية ومكافحته بعد مشاركتهم في البرنامج التدريبي، علاوة على ذلك، تم تطبيق المعرفة والمهارات المكتسبة بنجاح من قبل المشاركين، الذين قاموا بمشاركتها لاحقًا مع صانعي المحتوى الآخرين. وتقدم الدراسة عددا من المقترحات حول أفضل الممارسات لخلق بيئات إيجابية وشاملة على المنصات الرقمية، وتحسين برامج التربية الرقمية التي تهدف إلى تمكين صانعي المحتوى من مواجهة خطاب الكراهية عبر الإنترنت.

الكلمات المفتاحية: خطاب الكراهية، الأمية الرقمية، الإنترنت، برامج التدريب.

---

* Assistant-lecturer at the Faculty of Mass Communication, MSA University

Abstract

This paper examines the efficacy of digital literacy training programs in countering online hate speech. Through a qualitative investigation, the study delves into the experiences and perspectives of content creators who took part in the Radio Netherlands Training Centre's 'Digital Media to Counter Disinformation and Hate Speech Training Program' during September/October 2022. The study employed in-depth semi-structured interviews with a diverse sample of 10 participants hailing from different countries. The findings elucidate that the participants perceived a notable enhancement in their proficiency to identify and combat hate speech following their involvement in the training program. Moreover, the acquired knowledge and skills were successfully applied by the participants, who subsequently shared them with other content creators. Significantly, the participants themselves contributed recommendations regarding optimal strategies for fostering positive and inclusive online environments, as well as enhancing critical digital literacy programs that aim to empower content creators in their efforts to combat online hate speech.

Keywords: hate speech, digital literacy, training programs.

Introduction

The widespread availability of the Internet and social media has greatly expanded the opportunities for individuals to express themselves freely, diminishing the traditional media's role as gatekeepers. This is generally viewed as contributing to the democratization of public discourse. However, the limitless potential for free speech has also paved the way for the polarization of public opinions, aggressive forms of communication, and the detrimental effects of online aggression.

Surveys conducted globally have revealed that many internet users have reported being subjected to hate speech, and developing countries are not immune to its negative impacts, including psychological harm, marginalization, and exclusion [1]. Research has shown that individuals in developing countries face unique challenges when it comes to dealing with online hate speech, due in part to limited access to

technology, insufficient legal frameworks, and inadequate capacity to monitor and regulate the internet [2].

Online hate speech often takes the form of discriminatory language directed at marginalized communities, such as ethnic, religious, and linguistic minorities, as well as other vulnerable communities in significant ways. This has major impacts on the targeted individuals and communities, contributing to social divisions and exacerbating existing inequalities.

In response to the problem of online hate speech, researchers have called for a range of interventions. Some authors have highlighted the potential of technology, specifically the utilization of machine learning algorithms, to identify and detect hate speech on digital platforms [3]. Others have focused on strengthening the legal frameworks [4], while others have emphasized the importance of education and awareness-raising efforts aimed at promoting responsible and ethical online behavior [5].

Several initiatives have been undertaken to counter online hate speech and promote conscientious online behavior. One such initiative is digital literacy training programs for content creators. These programs are designed to educate individuals about the negative impact of hate speech and equip them with the skills and knowledge to identify and counteract online hate speech in their work and online interactions.

Despite the fact that these programs have been shown to have a positive impact on individuals' online behavior, including reducing online hate speech, there is limited research on the experiences and perspectives of content creators involved in these programs. This research aims to fill this gap in knowledge by exploring the experiences and perspectives of content creators involved in the Radio Netherlands Training Centre (RNTC) training program entitled *'Digital Media to counter disinformation and hate speech'*.

Problem Statement

Online hate speech is a growing problem that undermines social cohesion and threatens the safety and well-being of individuals and communities. Despite increased efforts to address this issue, the persistence of online hate speech highlights the need for more effective procedures to counter its spread.

This study seeks to gain a deeper understanding of the ways in which digital literacy training programs can help content creators navigate the complex and rapidly evolving online environment and reduce the spread of online hate speech.

Research objective

The main objective of this research paper is to gain insight of the experiences and perspectives of content creators who have participated in the Radio Netherlands Training Centre (RNTC) training program entitled 'Digital Media to counter disinformation and hate speech' in September/October 2022 round, and to examine the impact of digital literacy training programs on countering online hate speech and promoting responsible online behavior.

Literature Review

This literature review examines the existing research on online hate speech, including its definition, causes, consequences on targeted groups, and current responses. It also tackles the research that examined the role of critical digital literacy to counter hate speech, giving an overview of three digital literacy training programs targeted at content creators to empower them to combat hateful online content.

Online Hate Speech; a Growing Phenomenon

As the internet becomes more mature, the issue of online hate speech has gained more attention, especially as its offline consequences become more apparent. This type of speech ranges from misogyny and racism to xenophobia, blasphemy, and homophobia.

'Hate speech' is an emotionally charged term, with no universally agreed-upon definition in international human rights law. Despite many individuals claiming to recognize it when they encounter it, the criteria for identifying what constitutes "hateful" is often unclear or inconsistent, and invites differing legal, ethical, and socio-technical perspectives [6].

Several definitions of "hate speech" have been created in response to discriminatory incidents and social issues. These definitions have evolved

to address new situations, changes in language, increased understanding of equality and discrimination, and advancements in technology.

In the context of this paper, the term hate speech is understood as any purposeful attack directed towards an individual or a group, motivated by the targeted entity's identity or opinions. According to the Encyclopedia of the American Constitution, the identity factors include race, religion, ethnic origin, national origin, sex, disability, sexual orientation, or gender identity [7].

Rather than prohibiting hate speech in general, international law prohibits the incitement to discrimination, hostility, and violence. If hate speech doesn't reach the level of incitement, international law doesn't require countries to prohibit it.

One of the primary drivers of online hate speech is the anonymity provided by the internet. According to a study by Brink et al. (2019) [8], the anonymity of online interactions allows individuals to express hateful views and engage in abusive behavior that they might not otherwise exhibit in face-to-face settings. This anonymity also creates a sense of distance and detachment from the harm caused by hate speech, further exacerbating the problem.

The structure and algorithms of social media platforms have also been identified as contributing factors to the spread of online hate speech. The algorithms that determine which content is promoted and visible to users can amplify hateful and extreme views, creating echo chambers and reinforcing the views of those who engage in hate speech.

According to Statista 2021, the majority of cases of hate speech target individuals based on their political views (55%), followed by belonging to a protected class

(35%)[*], and physical appearance (35%), then race and ethnicity (25%), and gender (25%), followed by religion (21%), occupation (12%), disability (11%), and sexual orientation (9%). HateBase, a web-based

---

[*]The protected classes include: age, ancestry, color, disability, ethnicity, gender, gender identity or expression, genetic information, HIV/AIDS status, military status, national origin, pregnancy, race, religion, sex, sexual orientation, or veteran status, or any other bases under the law.

application that collects instances of hate speech online worldwide confirms that incitements of hatred focusing on religion and class have also been on the rise since 2015 [9].

The effects of online hate speech are significant and far-reaching. A study by Jackson and Nafus found that exposure to online hate speech can lead to increased levels of anxiety and depression, as well as a decrease in self-esteem and sense of belonging for the targets [10]. Additionally, the spread of hate can contribute to the normalization of hateful and bigoted views and beliefs, in addition to the spread of misinformation and harmful conspiracy theories [11].

Social isolation during the COVID-19 pandemic, and the related elevation in online activity, have increased young people's vulnerability to hate-fueled narratives spread through social media and messaging apps. They can either be direct targets of hate speech or merely observers susceptible to influence [12]. In a 2022 study focused on Algerian youth, it was discovered that social networking sites have evolved into platforms where perilous conversations take place, thereby posing a significant risk to the well-being of young individuals [13].

Another Algerian study on the students of the departments of Psychology and Sociology at the University of M'sila found that the hate spread through social media has negative effects on the level of the social value system among youth and leads to violence. It also contributes to extremism and religious disharmony and weakens social cohesion [14]. The researcher emphasized the importance of diligently raising awareness about hate speech and racism, aiming to enable individuals to recognize and avoid such behaviors.

A study conducted with a group of Jordanian journalists highlighted the crucial recommendation of promoting awareness regarding hate speech and strategies to counteract detrimental online content. The study emphasizes the significance of discouraging social media users from sharing misinformation or manipulating facts without verifying their accuracy. Furthermore, it urges individuals to refrain from promoting extreme ideologies and concepts. [15].

Based on a study conducted with Egyptian youth, it was determined that hate speech is amplified within society due to the anonymity afforded to individuals who propagate hatred, violence, and racism through social networks. The study also identified the ease with which verbal aggression can be disseminated via comments and messages, as well as the convenience of inciting others without direct confrontation. Nevertheless, the research revealed a positive aspect wherein digital media aids young people reject extremism, intolerance, and hate by sharing stories of victims of terrorism and extremist organizations, showcasing the destruction caused by such groups, and promoting values of tolerance and moderation through community engagement and the production of a new, alternative discourse [16].

Social media researchers have noted an alarming increase in online sentiments characterized by anti-Chinese and anti-Asian prejudice. These sentiments can be attributed to fears associated with the coronavirus pandemic, with derogatory slurs and terms such as "kungflu" connecting the outbreak with specific ethnicities and nationalities. This surge in online animosity and mistrust has not only fueled virtual hate but has also translated into real-world consequences. Instances of verbal assaults, acts of violence, and boycotts targeting Asian-owned businesses have been on the rise as well. [17]. Conversely, several videos circulating within African communities feature black Americans offering cautionary advice to their counterparts in Africa, urging them to exercise extreme caution and protect themselves against viruses originating from the "white man."

Digital spaces also attack women with many forms of hate, including sexism, misogyny, and harassment, in response to their online activism and participation. An article by Fladmoe, A. & Nadim, M. [18] has demonstrated that online harassment can act as a silencing mechanism against women, causing them to exercise caution when expressing their views.

A recent study by the Pew Research Center have shown that one in two women suffer some form of gender-based online harassment, with percentages increasing in younger women under 35 [19]. A study pointed at Twitter, in specific, to be hostile platform for women, as

approximately one in ten tweets aimed at or about female individuals in the sample contain insults or other derogatory language [20].

The primary responses to online hate speech have mostly centered on legal measures, but this approach has its own risks and limitations. A narrow emphasis on legal remedies fails to consider how societies develop through disagreements and debates. Post, R., Hare, I. & Weinstein, J. claim that despite hate speech being a hurtful form of dissent, it can also shed light on deep-seated tensions and inequalities that need to be addressed beyond just speech and online issues [21]. At the same time, "only the most egregious forms of hate speech, namely those constituting incitement to discrimination, hostility, and violence, are generally considered unlawful." [22]

For that, the United Nations set a strategy in 2019 to fight online hate speech; the strategy and its execution should align with the right to freedom of expression and opinion, as it believes that more speech, not less, is the key to addressing hate speech. The UN is also committed to empowering a new generation of digital citizens to recognize, resist, and combat hate speech [23].

Intermediaries are also continuously working on establishing rules for the appropriate use of their services and reserving to themselves, in specific situations, the role of moderators. These "terms and conditions," sometimes referred to as "community standards" or "guidelines," differ in the types of expression they restrict, with many prohibiting "hate speech" or similar terms. A yearly assessment implemented by POLITICO[†] on the handling of online hateful posts by Facebook, Google, and Youtube, revealed that 62.5% of flagged material was deleted by social media companies over a 6-week period between March 1st and April 15th, 2021. This is a decrease from the 71% removal rate recorded during the last review conducted in late 2019.

In a contrasting observation, the audit conducted by the Commission revealed that both Twitter and Instagram witnessed a rise in the removal

---

[†] A German-owned political newspaper company based in Arlington County, Virginia, US, that covers politics and policy in the United States and internationally.

of hate speech content subsequent to receiving notifications. Notably, TikTok demonstrated the highest removal rate, reaching 80%, among the prominent platforms that voluntarily underwent the review in 2022.[24]. Furthermore, reports published in December 2022 by the Center for Countering Digital Hate and the Anti-Defamation League indicate a significant increase in the prevalence of hate speech on Twitter under Musk's leadership. [25].

In general, online hate speech is a complicated and expanding problem that demands a comprehensive approach. Although there is no silver bullet, studies have revealed that a combination of regulation, education, raising awareness among media professionals and citizens and technology can help curb the dissemination of online hate speech and mitigate its harmful effects.

Critical Digital Literacy (CDL) To Counter Hate Speech

Research has shown that hate speech is a pervasive problem on social media platforms, with the anonymity and lack of accountability afforded by these platforms enabling users to make inflammatory and harmful statements without consequences. However, it has also shown that critical digital literacy can play a role in counteracting hate speech by empowering individuals to identify and reject such speech, as well as by promoting a culture of responsible and ethical online communication.

Critical digital literacy (CDL) entails a digital literacy approach that places substantial emphasis on critically analyzing the power dynamics inherent in digital technologies [26]. This approach goes beyond mere proficiency in using digital tools, aiming to cultivate an understanding of how these technologies influence and are influenced by the broader societal context.

One key aspect of CDL is media literacy, which involves the ability to critically analyze and evaluate information found online. Studies have shown that individuals with higher levels of media literacy are more likely to protect themselves from harmful digital content and respect others [26]. The other important aspect of CDL is digital citizenship, which involves responsible and ethical behavior online. This can

include being aware of one's digital footprint, practicing good online etiquette, and understanding the potential consequences of one's actions online. Studies have shown that individuals with a strong sense of digital citizenship are less likely to engage in cyberbullying and hate speech [27].

In addition to media literacy and digital citizenship, CDL also encompasses the ability to navigate and utilize digital tools and resources. This entails understanding effective usage of search engines, proficiency in creating and editing digital content, and knowledge of safeguarding one's digital security.

Although there's a growing body of research on CDL to counter hate speech, such as one recent study by Foss SK, et al. et al. that found that individuals who had high levels of CDL were more likely to recognize and report hate speech when they encountered it online [28], existing studies predominantly focus on initiatives aimed at residents of the Global North [29]. For instance, Zeiger and Gyte highlight the EU-funded program called "No Hate Speech" as an example. This initiative has developed a comprehensive set of educational materials to teach media literacy and critical thinking skills, with a focus on identifying and countering hate speech online.

An emancipatory action research conducted with a group of Roma youth in Greece [26], which aimed to explore the impact of critical digital literacy activities on their understanding of online hate speech suggested that the most empowering aspect of the project was gaining knowledge and skills in creating non-violent counter-narratives. One the most interesting insights was that the participants considered their online presence and digital footprint in connection with their cultural identity, allowing them to see a larger connection between the online and offline social world. In other words, they linked hate speech to the larger issues of racism that they experience in the real world.

One report published by K4D Helpdesk‡ entitled 'Countering online misinformation, hate speech or extremist narratives in the Global

---

‡ The knowledge, evidence and learning for development (K4D) helpdesk is a service commissioned by the UK Department for International Development and other

South' report examined various initiatives from the Global South, with a specific emphasis on Pakistan [30]. The most prevalent type of initiative covered in the literature is countering violent extremism (CVE) programming, which encompasses not only countering extremist narratives and hate speech but also promoting digital literacy to address misinformation spread by extremist groups. The report provides a concise overview of the lessons learned concerning the best strategies to build the counter-narratives, including a precise identification of the target audience and tailoring messages to align with their specific attributes, as well as considering local cultural and contextual factors, and creating a safe online environment for minority groups who have been subjected to hate in the digital realm.

In an article by Isheba Cornwall, a Caribbean social media strategist, radio host, and producer, the author emphasized the effectiveness of creative campaigns via social media platforms in helping to combat online hate speech and toxicity. She suggested focusing on easily understandable content that promotes inclusiveness, equality, and diversity [31].

Another study implemented a randomized experiment involving English-speaking Twitter users who had previously posted messages containing xenophobic or racist hate speech. These users were divided into three groups, each assigned to a different counter-speech strategy. The study then analyzed the efficacy of each strategy. The findings revealed that counter-speech messages emphasizing empathy were successful in reducing instances of xenophobic hate speech. However, strategies utilizing humor or warnings of consequences did not consistently yield the same positive outcomes [32].

Furthermore, according to a report published by IDEAS Sri-Lanka (2019), fostering trained youth advocates for peace through social media can be advantageous in countries experiencing heightened levels of tension. The report suggests that empowering young individuals with a comprehensive understanding of the country's context, political

---

Government departments, that provides brief summaries of current research, evidence, and lessons learned.

structure, conflicts, and peace building, along with providing them with tools to effectively monitor various information sources, particularly during periods of violence, is crucial. Additionally, the report recommends promptly countering hate-inducing media messages following violent incidents as a means to cultivate peace and foster empathy among citizens [33].

Training content creators has also emerged as a promising approach to countering online hate speech. The literature suggests that training programs can equip individuals with the knowledge and skills needed to recognize and respond to hate speech, promote inclusive and respectful online environments, and counter the spread of harmful ideologies [34, 35].

In addition to educational interventions, technology plays a significant role in advancing and supporting critical digital literacy (CDL). For instance, a study conducted by Kim et al. revealed that the implementation of machine learning algorithms for the detection and flagging of hate speech can effectively contribute to CDL promotion and combating online hate speech [36]. The field of natural language processing (NLP) is continuously evolving in terms of utilizing AI techniques for preventing and intervening in online hate speech. Recent research has focused on evaluating the effectiveness of various models employed by companies and media organizations for this purpose [37]. Nevertheless, the primary focus of this paper is on the human element, specifically content creators in the realm of digital media.

Overall, the literature suggests that CDL is a valuable tool for counteracting hate speech online and that it can be effectively developed and promoted through educational interventions and the use of technology.

Overview of a Sample of Critical Digital Literacy Training Programs

In this section, the researcher provides an overview of three distinct short-term training programs focused on critical digital literacy (CDL) in which she participated: two as a participant and one as a trainer. These programs were conducted in three distinct countries and offered full or partial scholarships to ensure inclusivity across various socio-

economic backgrounds, as well as to accommodate students and young individuals. Despite slight variations, these programs shared similar objectives and specifically targeted content creators from developing nations.

Media and Digital Literacy Academy in Beirut (MDLAB)

Established in 2013, the Media and Digital Literacy Academy of Beirut (MDLAB) worked on the development of media and digital literacy education in the Arab region by delivering the necessary training, curricular material, and resources. MDLAB has been at the forefront of media and digital literacy initiatives in the Arab world, thanks to the country's relatively open media and academic environment, when compared to other countries in the region [38].

At the Lebanese American University in Beirut, Lebanon, MDLAB offers an intensive two-week training program held annually during the summer. The program's primary audience consists of faculty members, media professionals, content creators, and media students from around the globe, with a particular focus on individuals from developing countries. The program includes a series of lectures, panels, and practical workshops on digital competencies. It covers diverse topics such as media and digital literacy, propaganda in video and film, the dynamics of hate speech and free speech, the influence of digital media on radicalization, and the portrayal of marginalized communities in the media. The practical workshops delve into various skills such as blogging, podcasting, photo manipulation, audio and video editing, social network analysis, and data visualization. Upon completion of the program, academics are expected to develop a syllabus suitable for teaching at their respective universities or educational institutions, while content creators are tasked with producing content that promotes critical media and digital literacy.

RNTC's Digital Media to Counter Disinformation and Hate Speech Course

The Digital Media to Counter Disinformation and Hate Speech course is an intensive training program offered by the Radio Netherlands Training Centre (RNTC) that spans over three weeks. Introduced in 2015, this program specifically caters to content creators, journalists, social activists, program-makers, and communication professionals

from developing countries who aspire to drive social change within their communities, basically through producing creative digital media stories, to micro-target audiences.

A significant portion of the program is dedicated to enhancing participants' abilities to monitor, track, and combat hate speech, cancel culture, and polarizing narratives. The curriculum aims to empower potential agents of change by providing them with proactive strategies for addressing these issues. Additionally, the course educates participants on effective community moderation techniques to foster inclusive dialogue. Tools such as social listening, data verification, and understanding virality are imparted to aid in this endeavor. [39].

By the end of the program, participants are anticipated to produce online videos that influence attitudes and behaviors, applying what they learnt about various aspects such as shooting, lighting, audio, and video editing, as well as different story-telling formats.

American Center Cairo 'Digital Skills Camp'

The Digital Skills Camp is a program offered by the U.S. Embassy American Center Cairo (ACC). It was initiated by a team of media experts and educators during the summer of 2021. The program mainly aims to promote conscious and critical media and digital literacy among Egyptian youth. It caters to both established content creators and aspiring ones [40].

The two-week-long program educates participants on how to merge critical thinking, audio/video production skills, and social media moderation tools to produce meaningful media messages that could address pressing topics in the online space, such as fake news, hateful and discriminatory speech, extremism, and cyberbullying, and put forward responsible and ethical content. Audio and video editing workshops are also an integral part of the training. One plus point of this program is that some lectures are open to public audiences for greater benefit.

Like the two other programs, participants are required to bring what they learned in the training program to practice. The produced content should be a podcast or a video content tackling an issue that can be

solved by educating audiences to critically think and skillfully engage with the media content they're exposed to.

A common denominator of the three initiatives is the emphasis on the development of critical thinking skills and the ethically-reflective use of social media as starting points of media and information literacy skills to combat hate speech online. In this paper, the researcher specifically chose to examine the experience of RNTC as a case study due to its diverse participant base in terms of nationalities, occupations, and years of experience. Another reason for selecting RNTC is that the program concluded last October, so the experience should be still vivid in the participants' minds. Additionally, previous academic papers and articles have already covered MDLAB, such as the paper by Jad Melki and Lubna Maaliki titled "Helping Arab Digital and Media Literacy Blossom: Three years of the Media and Digital Literacy Academy of Beirut (MDLAB)"[41].

Therefore, the following research questions were formed:

RQ1: How do content creators perceive and assess the effectiveness of the RNTC program; *Digital media to counter disinformation and hate speech* on their knowledge and skills?

RQ2: Which learning techniques proved to be the most effective in enhancing the participants' knowledge and skills?

RQ3: How did content creators apply the knowledge and skills they obtained from the training in their work and other online activities?

RQ4: Were there any efforts from the participants to diffuse the training materials to other content creators?

RQ5: How can content creators counter hate speech and contribute to creating a safe and inclusive online environment?

RQ6: How can digital literacy training programs be specifically designed and implemented for content creators to more effectively reduce online hate speech?

Methodological procedures

Research design

This study adopts a qualitative research approach that focuses on how individuals interpret and make sense of their experiences to grasp social reality. It is inductive in nature, and the researcher generally explores meanings and insights in a given situation [42], by utilizing various methods such as interviews, diaries, classroom observations, and open-ended questionnaires to gather information and analyze it through content analysis of visual and textual materials and oral history [43]. It aims to provide a deeper understanding of human behavior, emotions, attitudes, and experiences. The research is exploratory, seeking to explain the "how" and "why" of a social phenomenon in a specific context [44].

The use of qualitative methods is appropriate for this study, as it can provide rich, detailed data that can provide insights into the lived experiences of content creators and their perspectives on their participation in the RNTC training program 'Digital Media to counter disinformation and hate speech, their perceived benefits and drawbacks of the training program, their opinions on the program's content and delivery, and its impact on their work and online behavior.

Instrumentation

The researcher conducted a ***snapshot case study***, where a case is being examined in one particular period of time [45], such as a current event, a daily diary, etc. The analysis is facilitated by observing events within a limited time frame, whether it be a day, week, month, or even an hour. The final result is a complete picture of the situation within a condensed period of time.

This particular case study is well-suited for this paper due to the duration of the RNTC training program, which lasted for a concise three-week period. The researcher's objective is to explore the impact of this relatively short timeframe on the knowledge, skills, and expertise of the content creators involved.

Data collection was conducted through ***in-depth semi-structured interviews*** with content creators who participated in the RNTC digital

literacy training program. The interviews were designed to delve into their firsthand experiences of the training, covering aspects such as the content covered, delivery methods employed, and their personal assessments of how the program influenced their capacity to combat online hate speech.

The semi-structured interviews data was analyzed using qualitative data analysis techniques, including content analysis and thematic analysis, to identify patterns and themes in the participants' experiences and perspectives.

Semi-structured interviews are suitable for this study as they provide a flexible structure for exploring the experiences and perspectives of the participants [46].

The interview guide was developed based on a review of the literature on critical digital literacy training programs and online hate speech, and was comprised of eight open ended questions to elicit detailed responses from participants.

Data collection procedures

Eight interviews were conducted via zoom and lasted between 30-35 minutes. Due to different time zones and working hours, two participants preferred to get the questions sent to them by email, and answers were sent back later via an audio app.

The interviews took place three months following the conclusion of the training program. All participants were fluent in English language, thus enabling the majority of the interviews to be conducted in English.

Data collection was conducted in a systematic and ethical manner, with informed consent obtained from all participants to record the interviews.

The interviews took around five hours on four different days. The recorded interviews were transcribed using verbatim, over a period of two days before content analysis was done to generate the themes.

Participants

The sample for this study consisted of content creators who had participated in the RNTC training program 'Digital media to counter disinformation and hate speech'. The participants were recruited through **purposive sampling**, in which individuals are selected based on specific criteria relevant to the research question. In this paper, the criteria included attending the RNTC training program in person during the September/October 2022 round, having years of experience in creating content online, in addition willingness and availability to participate in the study.

Out of the 16 participants (excluding the researcher) who took part in the RNTC training program, the researcher contacted 13 individuals. Among them, a total of 10 participants (equivalent to 63% of the participants) agreed to be interviewed.

Demographic characteristics of the participants

Participant	Name	Country	Occupation	YOE
#1	Atiol ElMalik	South Sudan	Journalist, photographer & film maker	19
#2	Aysheh Alghabin	Jordan	News reporter, freelance investigative reporter with a long experience as a social media specialist and content officer	11
#3	Diaa AlYassin	Syria	Investigative journalist & political economy researcher	8
#4	Elias Ahamed Khan	Bangladesh	Broadcast journalist and digital content creator	7
#5	Manal AlHamshri	Libya	Reporter & media monitor of political pluralism in Libyan channels	4
#6	Menna Ayman	Egypt	Senior copywriter & freelance journalist	7
#7	Nombre Souaibou	Burkina Faso	Journalist & editor	10

#8	Nour Safieddine	Lebanon	Senior media projects coordinator, freelance journalist & a religious freedom certified trainer	10
#9	Phương An	Vietnam	Broadcast journalist & content creator	7
#10	Rachel Yusuph Mariam	Tanzania	Assistant lecturer & radio presenter	14

Measurements

A set of questions were prepared by the researcher to cover the following areas:

➢ Participants' acquired knowledge and skills from the RNTC training program 'Digital media to counter disinformation and hate speech digital literacy.

➢ Participants' views on the challenges and opportunities encountered when applying the knowledge and skills gained from the training program in countering online hate speech.

➢ Identification of the most effective learning techniques experienced by the participants during the training program.

➢ Participants' efforts to share knowledge and skills with content creators.

➢ Recommendations from participants on the best practices to counter online hate speech and improve digital literacy training programs for the same purpose.

Results

Participants' Acquired Knowledge and Skills after Attending the RNTC Training

In this section, the researcher aimed to evaluate the perceived effectiveness of the RNTC program 'Digital media to counter disinformation and hate speech' by soliciting feedback from participants. The inquiry focused on the knowledge and skills they gained, with particular emphasis on identifying their most valuable takeaways from the program.

The audience mapping and storytelling recipe taught in the training course were unanimously praised by all participants as the most beneficial takeaways. The RNTC audience mapping techniques, designed to help content creators producing content for behavioral change, were particularly effective. By identifying attitude, behavior, skill, or knowledge gaps in the target audience, uncovering their beliefs, emotional motivators, and influencers, the mapping technique allows content creators to develop persuasive and emotionally appealing messaging.

Participants found the RNTC audience map easy to use and understand, but also provides detailed and thorough information about the target audience, which help them create influential messages that are relevant, credible, and emotionally compelling, which ultimately leads to better engagement and positive outcomes.

Manal, a reporter and media monitor from Libya, expressed that the RNTC training program had a significant impact on her ability to think critically and creatively about her audience. It introduced her to new research techniques that allowed her to gain a better understanding of her audience. "*I learned to consider not just their demographics, but also their lives, values, attitudes, beliefs, future goals, and emotional payoffs in order to create a powerful media message that resonates with them.*"

Nour, a senior media coordinator and a religious freedom trainer, from Lebanon, found the persuasion aspect of the training interesting, stating, "*I believe in the power of emotions in changing one's existing beliefs and motivating them to acquire new knowledge, skills, or attitudes. Additionally, I agree that relying solely on logic may not be sufficient to convince audiences.*"

Another participant found the RNTC map's approach to sensitive matters such as online hate speech to be particularly clever. "*The RNTC map takes a clever approach by utilizing emotional motivators of the audience to effectively reshape their existing perceptions*", Elias, a broadcast journalist from Bangladesh remarked.

As for the RNTC storytelling recipe, participants found it to be suitable for different content formats they produce, such as news reports, media campaigns, drama, and in-depth reports. They appreciated how the

recipe follows the fundamental structure of a story, effectively engaging the audience by building tension and leaving them eager to uncover what unfolds next.

Ann, a broadcast journalist from Vietnam who primarily works on news content, was pleasantly surprised to discover that the RNTC storytelling recipe was highly effective when applied to news reports. "*By following the RNTC formula, I can take my audience on a journey that builds tension, creates emotional connections, and leaves a lasting impression*," Ann commented.

Incorporating storytelling elements into investigative stories was an interesting point mentioned by Diaa, an investigative journalist and researcher from Syria, as it keeps the audience interested and invested in the story, help simplify the complex information, and attract wider audience. He also added, "*Proposing a central question, creating tension and character development, and then finding an answer for the question can help the audience understand the significance of the information we have uncovered.*"

The majority of participants also reported that they gained proficiency in recognizing hate speech, assessing its effects, and creating counter-arguments and alternative arguments to combat it.

Counter-narratives mainly focus on deconstructing existing hateful arguments, undermining their credibility, and re-characterizing those portrayed as enemies, while alternative narratives aim to replace hate-promoting narratives with positive messages that foster peace, unity, and coexistence.

Menna, a senior copywriter & freelance journalist from Egypt, believed that the training helped her to skillfully utilize fact-checking tools to identify and expose any false or deceptive information found in the discriminatory and misleading discourse of propagandists. "*In my final project, I mixed both factual and sarcastic counter-narratives to highlight absurdity or hypocrisy of certain arguments or viewpoints.*"

Rachel, an assistant lecturer and radio presenter from Tanzania, highlighted the influential role of humor in crafting counter-narratives against hateful messages, which she was initially introduced to through

the RNTC training course. However, she emphasized the need for cultural and religious sensitivity when employing humor to address significant issues, so as not to offend anyone. Rachel shared an incident where a video, intended to combat religious extremism using a humorous approach, was shown to the participants, leading to some Muslims perceiving it as offensive.

Nour and Diaa shared the belief that under difficult circumstances, alternative narratives that aim to promote understanding and empathy among people are particularly impactful.

Nevertheless, half of the participants expressed that the topic of hate speech wasn't allotted sufficient time during the training. They proposed dedicating additional time to employing the techniques taught to influence individuals and instigate change regarding subjects that are associated with hate speech.

Participants also highlighted their significant acquisition of knowledge in social media moderation tools, particularly emphasizing their grasp of social listening and a comprehensive understanding of amplification and virality.

"*Online norms around certain topics might be extremely different than what we think they are. Social listening tools can help us to combat hateful sentiments in a timely manner to prevent them from escalating and potentially averting future crises*," Aysheh, news reporter and freelance investigative journalist from Jordan said. Rachel agreed with Aysheh, describing the 'social listening' as an '*impressive*' process of detecting emerging trends and stories before working on countering them.

For storytellers like Aitol, a journalist, photographer and film maker from South Sudan, amplification is a critical aspect of their work, therefore, obtaining knowledge about it during the training was something he cherished. Aitol expressed, "*Learning about how algorithms work has allowed me to comprehend why certain online content is given priority on social media due to their potential to spread and normalize harmful views. However, on the other hand, I've also gained insight into how algorithms can identify and oppose hate speech,*

as well as elevate narratives that endorse diversity, inclusivity, and human rights."

Elias shares Atiol's view that the trainers provided comprehensive coverage of the topic of amplification.. They discussed how a single comment or post can rapidly escalate into a wave of hatred that causes significant harm, which can spill over from the online to the offline world. Elias emphasized that this is particularly relevant in communities with political and religious tensions and tribalism.

For Ann, knowing how each social media platform works and what type of content is best suited for each platform is a key to success imparted by trainers to trainees. She stated, "*This includes understanding the anatomy of successful stories on each platform, the best factors to include in posts, and what make a content go viral.*"

Two respondents also mentioned gaining some knowledge about legal and policy frameworks related to hate speech, internet intermediaries' guidelines to combat hateful content, and online privacy and security.

Regarding the skills gained through the training, there were differing views on the workshops for photography, filming, audio, and visual editing. Those who had no prior experience in these areas found them helpful, while those who were already skilled in producing digital content found the workshops to be too basic and not helpful in enhancing their skills.

Nombre, a print journalist and editor from Burkina Faso, is used to relying on words as his primary means of storytelling. However, he recognizes that working in the digital age demands a broader skillset. "*Learning to shoot and edit videos was unfamiliar to me, and I encountered some difficulties in acquiring certain techniques, but with practice, it became easier,*" he said. Rachel, on the other hand, was delighted to have used the skills she learned in the training to create video content for the first time. However, both Nombre and Rachel wished that the practical workshops had been given more time.

In contrast, the workshops didn't prove to be very useful for the group of broadcast journalists and digital content creators. However, Elias identified a significant takeaway from the training, which emphasized

the importance of editing in shaping the final product and how to use mobile devices proficiently to create high-quality video content while adhering to ethical standards.

Moreover, the majority of participants believed that the training equipped them with verification kits or tools that can further enhance their skills in verifying information and data, hence assessing the reliability and accuracy of digital media content. These tools typically include software programs, online resources, and reference materials.

Most Effective Learning Techniques

The following section of the paper focuses on the perspectives of participants regarding the most effective learning methods employed during the training program. Participants agreed that group exercises were the most beneficial among the various learning techniques used by the RNTC trainers to convey knowledge and skills to combat online hate speech, followed by presentations, open discussions, and demonstrations respectively.

"*Creating working groups and allowing them to brainstorm, reflect and analyze various media content, formats, and techniques before presenting their findings was highly beneficial,*" Nour explained, "*The process could be difficult at times due to the diverse backgrounds of the participants, but working as a team promoted collaboration, enhanced participants 'communication skills, and led to more innovative and creative outcomes while also facilitating peer learning.*"

Rachel believed that activities helped to create a sense of belonging and shared purpose among participants, she added, "*Even when some clashes happened, I learned to be patient, present my viewpoint in a positive manner to my group. Alternatively, I allowed them to recognize the validity of my perspective through the feedback we received from our peers during our work presentations. These skills have proven immensely valuable to me in my work with students.*"

Elias, on the other hand, noted that group exercises offered him a chance to network and build relationships with professionals from varying backgrounds, "*These connections can lead to potential future collaborations and opportunities,*" he confirmed.

Based on the feedback provided by the participants, it appears that the majority found the presentations conducted by the trainers to be effective. Specifically, seven out of ten participants reported that the presentations were clear, engaging, and contained compelling visual aids.

The fact that a majority of participants found the presentations effective is a positive sign, as it suggests that the trainers were able to successfully communicate their message and engage their audience. Furthermore, the use of compelling visual aids can be particularly good in capturing and maintaining the attention of the trainees, which can help ensure that the message is received and retained.

The majority of participants, specifically eight out of ten mentioned they felt comfortable to ask questions and intervene during the presentations. This approach to training is known as interactive learning, where trainees are active learners, not passive recipients.

Regarding open discussions during the training, half of the participants found them to be fruitful and engaging. However, the other half felt that the discussions did not meet their expectations.

Ann appreciated the diversity among the participants, recognizing that it enriched the discussions. She remarked, *"Although we come from different worlds and had distinct perspectives on various situations, we still showed each other respect and understanding."* Aysheh affirmed Ann's statement by saying, *" some participants may have held biases, there's no doubt that the open discussions brought positive communication that helped overcome prejudice and bias."*

Menna also shared her thoughts on the diversity of the participants, stating that it was an eye-opening experience for many as they were exposed to seemingly irrelevant and far-fetched opinions. She emphasized, *"It served as a valuable reminder that people's backgrounds and ethnicities attribute a lot to their perspectives and opinions."*

Conversely, Diaa criticized the moderation of discussions during the training. He believed that the participants were not provided sufficient time or opportunities to discuss incidents of online hate in their respective countries or the initiatives undertaken to combat it. Diaa stated, *"I am confident that Arab countries have produced campaigns*

that are worth discussing and can be even better than those produced in developed countries."

Nour suggested that the training program should include an overview of the media landscape in each represented country. According to her, "*Having some background knowledge about the media landscape in each country would have been helpful. It would have allowed us to better understand the challenges and opportunities in addressing online hate speech, find common ground among ourselves, and learn from each other's experiences.*" This recommendation emphasizes the significance of contextual understanding when addressing complex issues such as online hate speech.

Furthermore, Nour stressed the importance of improved time management during the training. She stated, "*Better time management would have ensured that everyone had an equal opportunity to express their views.*"

Overall, both Nour and Diaa's comments highlight the necessity of adopting a comprehensive approach to training that takes into account the diverse perspectives and experiences of each participant. For them, incorporating contextual knowledge and promoting inclusive discussions can help create a more meaningful and effective learning experience.

Additionally, while demonstrations were not as prominently recalled by the participants compared to other learning techniques, Atiol specifically highlighted this method as a significant aspect of the training. He explained, "*Even if someone is already proficient and knowledgeable in a specific field, there are always fresh insights to be gained from experts with different experiences.*"

Participants' application of acquired knowledge and skills

In this section of the study, participants were invited to share specific instances where they applied the skills and knowledge acquired from the RNTC training to intervene or counter online hate speech in their work and online activities.

Despite the particularities of the content and the audiences they address, all participants acknowledged their ability to apply the knowledge and skills gained from the training course. They mainly intervened through

four measures; raising awareness about hate speech and its consequences, media monitoring, responding to hate speech, and creating peaceful and inclusive digital content.

"Taking a stance against both online and offline hate speech is an integral part of my daily work," stated Nombre. He further elaborated on the challenging security situation faced by Burkina Faso since 2015, characterized by terrorist threats and attacks targeting military barracks, schools, health facilities, and other public entities. In such circumstances, Nombre explained, "*Some individuals or groups may resort to hate speech as a means to assign blame and demonize certain groups, thereby exacerbating tensions and potentially leading to conflicts.*" He emphasized their responsibility as media professionals to "*report on issues in a fair and unbiased manner, providing a platform for diverse perspectives and countering hate speech.*"

Elias pondered upon the predicament of religious minorities in Bangladesh, where they are subjected to persecution based on their religious beliefs. According to him, some Muslim leaders exploit their mainstream and social media presence to instigate the majority against the minority groups with hatred both online and offline. He remarked, "*They spread false information about a particular group or individual through news, photos, or even videos, which create a false narrative or belief that fuels animosity.*" This, in turn, "*leads to the formation of stereotypes, prejudice, and discrimination, and often escalates into harmful behaviors or attacks on the homes or businesses of minority groups.*"

Elias, along with other young content creators, are committed to constantly enhancing their knowledge and expertise in the field of digital media. As he stated, "*We strive to use social media to foster social dialogue, empowering people to become critical consumers of information and responsible sharers of knowledge.*" Elias is determined to ensure that individuals are able to practice their religion without any fear of persecution or discrimination.

Egyptian society has also been highly polarized in recent years, with deep political and ideological divisions that have spilled over into the online space. Menna believes that social media had contributed to the widening

gap by providing a platform for discussions that often lead to personal attacks, insults and harassment to those with different opinions.

As an active Twitter user, Menna uses the platform to engage in political debates and challenge hateful narratives based on either true or fabricated information or photos. She advocates for a civil and open-minded approach and says, "*Instead of responding with hateful replies, I focus on debunking haters' statements and arguments using the counter-narrative techniques I learned at RNTC. I correct their misinterpretations and validate their feelings while stating the truth as it is.*"

Additionally, Menna has used the techniques she acquired from the course, specifically mapping, social listening and verification to delve deeper into more relevant stories for her audience.

Both Aysheh and Rachel emphasized the issue of online harassment towards women, which they personally experience or observe in their respective situations. They highlighted how such hateful behavior have a chilling effect on women's participation in public discourse, as many women may feel unsafe or intimidated to speak out in the face of such offensive content.

According to Aysheh, a journalist with a focus on women's issues, Jordanian women are particularly vulnerable to online hate speech. She noted that they are often subjected to misogynistic and sexist language, with a disturbing tendency to blame the victims for the violence they endure. She stated, "*Whenever there is news about violence victims, I consistently encounter numerous hateful comments justifying violence against women.*"

In her efforts to combat hate speech, Aysheh not only closely monitors the comments and reports instances of hate, but she also emphasizes the importance of empowering women to respond to hate speech in a constructive manner. As she expressed, "*I always strive to encourage women to engage in dialogue, call out hate speech, and take concrete steps.*" By promoting active and assertive responses, Aysheh aims to foster a culture of standing up against hate speech while avoiding further escalation.

In a related context, Rachel recalled an incident that occurred a month after receiving the training, where a colleague shared fake news about Tanzanian men being the most faithful in the world. When she debunked the news, she was subjected to a barrage of online hate messages. As she explained, "*I got more than 100 messages that attack me for humiliating the men of my own country, making fun of me and shamed me for not being married. Some of the people who sent the messages were unfortunately my colleagues and other people I knew in person.*"

However, the training helped Rachel adopt a calm and rational approach, and she did not engage in hateful debates. She responded by informing the senders that it was her job to verify information and correct false news. She also stressed, "*I assured them I did not intend to offend anyone, and that exposing fake news had nothing to do with my personal choices. Finally, I let them know that their messages hurt my feelings.*"

Nour, a religious certified media trainer hailing from Lebanon, tirelessly dedicates herself to countering online and offline hate a country plagued by sectarian and political conflicts that perpetuate a culture of animosity and division. To address this issue, she actively moderates webinars on religious freedom, convening experts from diverse religious backgrounds. These webinars serve as platforms for discussing and emphasizing the significance of religious tolerance and acceptance in fostering harmony and unity among communities.

Moreover, Manal, as a media monitor of political pluralism in Libyan media channels, works to combat online hate in several ways, most importantly, reporting on instances of hate speech that occur in the traditional and digital media, bringing attention to the issue and highlighting the harmful impact of such speech. She also works to promote political pluralism in Libya, "*Me and my colleagues, make every effort to create an environment where diverse voices are heard and respected, reducing the likelihood of hate speech and promoting constructive dialogue.*"

As witnesses to the devastating effects of civil wars in South Sudan and Syria, Atiol and Diaa may be particularly aware of how online hate speech can contribute to the dehumanization of individuals and groups,

and how it can create an environment of fear and hostility. Their firsthand observations have revealed how various warring factions exploit social media platforms to rationalize acts of violence and suppression, ultimately undermining social cohesion and eroding trust within communities.

In 2013, South Sudan was engulfed in a cruel civil war, only two years after it gained independence in 2011. The country is still encountering significant obstacles in its quest for enduring peace and stability. The persistent violence and political unrest pose a grave threat to the nation's stability. As an entrepreneur in the field of photography and filmmaking, Atiol disclosed that he is venturing into diversifying his enterprise by developing digital content that advocates for inclusivity and tolerance and has the potential to reach a wider audience. "*I'm already incorporating the RNTC storytelling techniques in the multimedia stories I'm currently working on,*" he said.

As an investigative reporter, Diaa found great value in the emphasis that RNTC placed on the significance of fact-checking and data verification techniques when it comes to combating online hate speech. According to him, "*By building stories that challenge divisive and hateful narratives on verified facts and information, we can ensure that we do not inadvertently contribute to the spread of misinformation or false narratives.*"

Following the earthquakes that struck Syria in February 2023, there were efforts made by both the Syrian Regime and the opposition to garner sympathy for the areas under their control. However, these attempts frequently employed divisive language and fueled sectarianism. In response, Diaa and his colleagues from Syria utilized their social media platforms to advocate for a message of unity and empathy towards all those impacted by the earthquakes, irrespective of their political affiliations.

Diaa, currently residing in a refugee shelter in the Netherlands, shared his observation following the recent earthquakes in Syria and Turkey. He noticed that some Turkish refugees in the same shelter were posting racist content, making comparisons between the aid provided to Syrian and Turkish people. They even asserted that certain individuals were "more deserving of help" in the midst of a humanitarian crisis. Diaa

took action by reaching out to several non-governmental organizations, stating, "*I reached out to several non-governmental organizations that pledged to organize movements and awareness talks that promote acceptance and condemn hatred.*"

Diaa also expressed his view that RNTC missed an important point in their training. He believes that participants should be equipped with the necessary skills and resources to address urgent and immediate situations, especially when the journalist or content creator is personally affected. He stated, "*Sometimes, I'm hesitant to engage in hostile conversations to avoid getting too emotional or unintentionally offend others trying to counter-argue.*"

The anecdotes shared by the participants are a testament to the effectiveness of the RNTC training course in enhancing critical digital literacy skills to combat online hate speech.

Participants brought attention to the unique obstacles they face when confronting online hate speech in their individual nations, including the mistreatment of religious minorities, political and ideological divisions, and the targeted harassment of women online. Nonetheless, they also shared inspiring instances of utilizing the knowledge and abilities gained from the training program to combat hate speech and facilitate constructive conversations. Their dedication to promoting critical thinking and responsible information sharing represents a significant stride in curbing the dissemination of harmful content.

Participants' Diffusion of the Training Materials to Other Content Creators

A majority of the respondents, accounting for 60%, mentioned that they disseminated the materials from the RNTC training course to their colleagues, professors, students, and junior trainees, each adapting the resources to their respective contexts.

"*I'm proud to say that around 300 students who I taught last semester now have the basics of creating videos that can effectively combat harmful online content,*" As a university assistant lecturer, Rachel cherished the learning materials provided by RNTC that she could easily use in her classes.

Aysheh also stated that she has already shared the knowledge she gained from the training with her colleagues and they have begun to implement some of the media campaign strategies and storytelling techniques on their newspaper's social media platforms. She elaborated, "*The training was insightful, and I believe that sharing this knowledge with my colleagues will help us create more engaging content for our audience.*"

With his enhanced understanding and skillset, Atiol was eager to share his knowledge and help others produce quality digital content. Atiol's enthusiasm was evident as he shared, "*I am delighted that I can now mentor junior journalists and filmmakers to produce persuasive and impactful stories for digital platforms.*"

Ann has also recently been given an opportunity to train five of her coworkers in the team on the new skills she took away from the RNTC training. She also shared her experience with a national television magazine in Vietnam. Moreover, Elias shared what he learned with his faculty colleagues and professors. He stated, "*My class of 2022 were all introduced to RNTC storytelling techniques, hate speech concepts and correct measures to counter it, as well as online verification tools.*"

Nour's experience with sharing the materials was more cross border; as she created a document inspired by the RNTC storytelling formula for applicants of an EU religious freedom film grant from Sudan, Iraq, and Lebanon. She shared, "*I distributed this document among my Internews colleagues in the region, so they can use it for similar projects.*" The document asked candidates to formulate their main inquiry, generate suspense by depicting a sequence of incidents that the characters react to in a specific order, culminate in a peak point, and solve the issue or address the central query.

When participants of a training program diffuse or share the materials they have learned with other content creators, it is a step forward for several reasons; Firstly, it indicates that the participants found the training materials to be useful and valuable enough to pass along to others. This suggests that the training program was effective in providing relevant and practical knowledge and skills that can be applied in real-world contexts.

Secondly, the diffusion of training materials can have a ripple effect, as the content creators who receive the materials can also benefit from the knowledge and skills they provide. This can lead to a wider impact, as more individuals and organizations become equipped with the necessary tools to create high-quality content and combat issues such as online hate speech.

Thirdly, the diffusion of training materials can also foster a sense of community and collaboration among content creators. By sharing knowledge and skills, individuals can work together to achieve common goals and tackle shared challenges.

Participants' Recommendations for Best Procedures to Counter Online Hate Speech

Drawing on the participants' insights and experiences, this section presents a list of recommendations for best procedures content creators can take to counter online hate speech, and create a safe and inclusive online environment.

- **Promoting critical thinking skills among social media users, and providing education on the topic of hate speech.**

All participants agreed that critical thinking, education, and awareness are crucial elements in the fight against hate speech. "*Promoting critical thinking skills is key to combating hate speech*," Elias asserted. He suggested that digital media literacy programs are an effective way to teach individuals how to "*identify and respond to hate speech*" and "*develop tools to evaluate the credibility of online sources and recognize biased or misleading information.*"

In addition to promoting critical thinking, it is crucial to educate social media users about hate speech, as Nombre argued. He recommended providing information about "*what constitutes hate speech, the harm it causes, and how it can be reported,*" as well as teaching individuals about "*the importance of respecting diversity and promoting inclusion.*"

As Aysheh noted, "A*cknowledging hate speech and pointing it out is indeed an important step in countering it,*" stating that by bringing attention to harmful language and behavior, individuals and communities can educate people about the harm caused by hate

speech and encourage them to be more mindful of their own words and actions. "*Calling out hate speech can also send a message to those who engage in it that such behavior is unacceptable and will not be tolerated, contributing to a decrease in its prevalence*," she added.

- **Utilizing fact-checking as an essential weapon in tackling the propagation of hate speech in the digital era**

Another point of agreement among the participants was activating the fact-checking and verification tools as a defense against online toxicity. As mentioned beforehand in this paper, false information, misleading content, and biased news can fuel intolerance and bigotry, therefore Menna requested content creators to strive to correct misinformation, disinformation, and mal-information that take place online and offline. "*By verifying the accuracy of claims made in hateful content as well as the counter-hate content, we can prevent the spread of biased and misleading information as well as harmful views.*"

Nombre also emphasized on the importance of the dependability and impartiality of information sources, while also ensuring that the dissemination of this information does not cater to a particular agenda. Ann recommended that content creators should advise their audience to avoid relying solely on one information source and instead seek out multiple reputable media outlets. She explained, "*This would enable them to develop a more complete understanding of the topic and minimize the possibility of being misled by biased or false information.*"

- **Creating tailored counter/alternative narratives that cater to different target audiences**

Building on 'mapping techniques' and the 'social listening' previously discussed, eight out of 10 participants emphasized that effective communication is not a one-size-fits-all approach. Instead, it requires a careful analysis and understanding of the target audience, as well as the contexts in which they operate. "*It's important to note that messaging that works in one context may not necessarily work in another. That's why it's essential to understand and consider cultural norms, beliefs, and traditions that may impact how the audience*

receives and interprets messages, as well as their motivators to take action," Rachel stated.

She added that once the target audience has been identified, tailored messaging can be developed that is specifically designed to resonate with them. *"This means using language and terminology that is relevant and meaningful to the audience, and framing messages in a way that speaks to their values and concerns."*

Participants also stressed on the importance of employing social listening tools to gain insights into people's attitudes, opinions, and behaviors. *"By listening to the conversations and sentiments of the target audience, it is possible to gain a better understanding of the underlying issues and grievances that fuel hate speech,"* Manal commented.

Based on the participants' perspectives, it is understood that by recognizing this, counter-narratives can be crafted to address the concerns of the intended audience in a positive and constructive manner. These counter-narratives can be tailored to address the specific themes and arguments of the hate speech, while also offering alternative perspectives and solutions. By engaging with the target audience in this way, it is possible to shift their attitudes and beliefs, and ultimately reduce the spread of online hate speech.

- **Ensuring adequate representation for diverse communities and minorities, as well as creating a safe space for their participation on digital platforms.**

Given that many participants hailed from nations grappling with political, sectarian, or tribal conflicts, several suggestions were put forth regarding this issue. Elias and Manal stressed the significance of incorporating a variety of viewpoints and perspectives in public discussions and preventing the marginalization or exclusion of minority groups from important dialogues. Elias commented, *"Greater visibility of minorities in society can break down stereotypes, decrease prejudice and challenge the negative stereotypes that fuel hate speech."*

Nour added that when minority groups are more visible, they can showcase their multifaceted identities beyond the hate speech directed at them.

"*This serves as a reminder that minorities are not defined by the hate speech directed at them; rather, they are complete individuals with unique experiences and perspectives*," she explained.

Furthermore, they emphasized that better representation for minorities can help to create a sense of belonging and community, which can mitigate the effects of hate speech. "*When individuals feel that they are part of a larger community that values their contributions and perspectives, it can help to counter the feelings of isolation and vulnerability that are often associated with being targeted by hate speech*," Rachel added.

- **Reporting hateful content to social media platforms**

The majority of the participants acknowledged their responsibility to report any form of online hate speech and other types of content that breach the community guidelines of social media platforms. Aysheh emphasized the significance of being specific and accurate while reporting hate speech and using the right terminology to avoid normalizing or trivializing it. Aysheh explained, "*If you are reporting on racist content, it is better to use the term 'racist' instead of using general terms like 'hateful' or 'offensive,' and refrain from using stigmatizing language or reinforcing stereotypes about targeted groups.*"

Additionally, Nombre put forward the suggestion that content creators should take on the responsibility of reporting on the approaches adopted by social media platforms in addressing online hate speech. Nombre stated, "*By reporting on the different strategies of social media platforms in dealing with online hate speech, we can hold them accountable and push them to take more substantial measures.*"

- **Calling for harnessing the Power of Artificial intelligence (AI) to Combat Hate Speech**

Ann pointed out to the crucial role content creators can play in advocating and even practicing pressures social media platforms and media outlets for activating better algorithmic detection and flagging systems.

AI tools can be integrated into the strategies mentioned above to help combat hate speech and other problematic content online. For example, machine learning algorithms can be trained to detect patterns of hate speech and flag

it for review by human moderators. Natural language processing can be used to analyze the sentiment and context of text-based content, helping to identify hate speech and other problematic language.

Speech recognition technologies can also be integrated into the process, to analyze audio or video content for problematic language or gestures. Additionally, AI-powered chatbots can be deployed to engage with users and provide education on responsible reporting practices and the dangers of hate speech.

"By working together with journalists and content creators, AI can help to identify and remove hateful content more efficiently, allowing for more proactive and effective moderation," Ann commented.

- **Keeping a tolerating mind and avoiding emotional triggers or getting involved in meaningless disputes**

Nour and Menna discussed their compassionate approach towards combating online hate speech. Menna shared her belief in the power of nonviolent communication, saying, *"Words can be hurtful and can have a significant impact on people's emotions, and as such, I believe in validating other people's feelings to create a safe space for open communication."* She also recognizes the futility of engaging in silly feuds online, stating, *"It only leads to further hostility and negativity. Instead, we should try to understand and empathize with the other person's perspective and adopt a more peaceful approach."*

Nour's approach to combating online hate speech is best described in her own words: *"I believe in 'stopping' to listen to what the other person is saying, without being triggered by my own emotions or biases. By doing so, I can better understand where the other person is coming from and reflect on their needs, concerns, and thoughts."*

When Nour has listened attentively to the other person, she follows up by restating their words in her own language. She explains the significance behind this practice, stating, *"Repeating back what they've said in my own words shows that I have understood their perspective. It also validates their feelings and thoughts, creating a sense of trust and openness."*

Nour also emphasizes the importance of making requests instead of demands. She suggests, "*Instead of attacking someone for their views or actions, one should ask questions or make requests that can help them understand the other person's perspective better.*" Moreover, Nour highlights the value of focusing on areas of agreement rather than disagreement.

Both participants embody a humanized approach that values empathy and nonviolent communication. They firmly believe that by embracing this approach, they can play a significant role in establishing a safer and more inclusive online community where every voice is acknowledged and respected.

- **Presenting Alternative Narratives is More Effective than Countering Existing Narratives When Combatting Online Hate Speech**

While discussing strategies to combat online hate speech, three participants stressed on a noteworthy idea, which is favoring presenting alternative narratives over countering narratives. Diaa, in particular, explained the reasons behind this preference. According to him, "*Countering existing narratives can sometimes backfire and reinforce them by drawing more attention to them. Instead, presenting an alternative narrative can shift the focus away from the existing narrative and create a more constructive conversation.*"

Diaa also explains that countering existing narratives can sometimes lead to people become more entrenched in their existing beliefs. By contrast, presenting an alternative narrative can be more persuasive in changing people's minds and attitudes, as it offers a new and more positive perspective that people may be more open to considering.

Overall, the recommendations presented in this section of the study cover a range of strategies, including preventative measures, such as education and awareness-raising campaigns, as well as more reactive measures, such as monitoring and reporting systems.

By sharing these recommendations, the researcher aims to provide valuable insights to researchers, policymakers, and online platform providers seeking to develop more effective methods for countering online hate speech.

Participants' Recommendations for Improving Digital Literacy Training Programs to Better Support Content Creators Counter Online Hate Speech

Based on their experience in the RNTC training, as well as other local and international trainings they had attended, participants were asked to provide their suggestions on how to enhance the design and organization of media and digital literacy training programs. These suggestions aim to enhance the efficacy of such programs in helping content creators combat online hate speech and other forms of online toxicity.

- **Providing Trainees With Solid Theoretical Foundation**

Nour suggested that even though digital literacy trainings are mostly geared towards professionals rather than researchers or academics, they should still offer a strong theoretical basis to their participants, which she found lacking in the RNTC Digital Media to Counter Hate Speech training. She emphasized that theories and frameworks are essential for developing a comprehensive understanding of the complexities surrounding online hate speech and can inform effective strategies and interventions for reducing its impact. She stated, "*For me, people working in the media field should be impacted by people doing research.*"

Furthermore, Nour acknowledged that certain workshops incorporate interactive elements and gamification to enhance the training experience. However, she cautioned that "this approach may not effectively motivate individuals to take tangible action." Additionally, she highlighted that trainings tackling the topic of hate speech should focus on differentiating it from freedom of speech, which is an ongoing debate in many countries around the world. This was a significant oversight, as hate speech laws vary across countries and understanding this boundary is essential.

- **Avoiding Treating Developing Countries as a Monolithic Group**

During the discussion, all participants reached a consensus on the tendency of developed countries to generalize and treat all developing countries as a homogeneous group. This approach was deemed inadequate, especially in the context of training programs offered to developing nations. The participants agreed that each developing country possesses its unique economic, social, and political context and cannot be treated as a uniform entity.

Therefore, it is crucial to consider the specific needs and requirements of each developing country in designing and delivering effective training programs. Reflecting on the RNTC experience, Nour said, "*Participants who came from developing countries may have similar economic backgrounds, but they certainly have varying exposure, cultural influences, and media landscapes that should be taken into account when designing and delivering training programs.*"

Nour, Diaa, and Aysheh also highlighted the significance of using culturally-relevant examples to engage participants and enhance the effectiveness of the training. Aysheh criticized the RNTC training for not adequately addressing the impact of hate speech on targeted groups, particularly Muslims and Palestinians. Nour expressed her hope for more relevant and up-to-date examples to be used, while Diaa emphasized the importance of providing space for participants to share their challenging experiences with both online and offline hate, especially in light of the significant political and economic turmoil experienced in the MENA region over the past 13 years.

Additionally, Rachel and Diaa emphasized the importance of an initial stage that facilitates introductions among trainers and participants, as well as among the participants themselves. Rachel shared her initial hesitance in opening up due to the fear of being misunderstood, stating, "*However, afterwards I started to open up with other participants, shared experiences, and exchanged opinions.*" Diaa further elaborated on this point, by saying "*This phase is particularly crucial for individuals from developing countries who may not be accustomed to practicing free media or free speech, and who may feel hesitant to express their thoughts in a public setting.*"

- **Recognizing the Cultural Context of the Participants is Crucial for Trainers**

The participants in the study noted that they felt a stronger connection to trainers who were originally from developing countries during the RNTC training. While they recognized the expertise and skills of other trainers, they found those from developing countries to be more adept at providing relevant examples and grasping the contextual nuances, particularly in addressing online hate speech.

Some participants even advocated for digital literacy training programs to be led by experts from developing countries who possess a deeper understanding of the local context and the root causes of online hate speech. Nombre expressed his preference, stating, "*Trainers from developing countries are more likely to have firsthand experience with the issues being discussed, enabling them to offer insights and solutions that resonate with the local context.*"

However, participants also acknowledged the diverse expertise of trainers who came from both developed and developing countries.

- **Aligning Training Process And With Trainees' Needs And Expectations**

Diaa advocated for a collaborative approach to decision-making between organizations providing digital literacy training and the trainees themselves, emphasizing the importance of fostering a sense of ownership and active participation among the trainees. He explained, "*When participants feel involved in the planning process, including determining the areas of focus, desired knowledge and skills, pace of delivery, and other aspects of the training, they are more likely to take an active role in the training and fully engage with the material.*"

One of the justifications he provided for his suggestion was the variation in participants' experience levels and job roles. Diaa highlighted that the work dynamics of an investigative journalist, for instance, are significantly distinct from those of a marketer. Taking into account the participants' diverse experiences and professional objectives, Diaa proposed grouping them accordingly. He explained, "*By carefully analyzing the participants' backgrounds and career*

goals and forming groups based on that, the training content can be customized to address their specific requirements and address the challenges they face."

- **Enhancing Digital Literacy Programs Through Advanced Labs, Equipment, and Ample Training Time**

Atiol elaborated on the importance of having access to the latest software and equipment to create high-quality digital content. He cited examples such as, video editing software, graphic design software, animation tools, and other related software and hardware equipment. He said, "*Having advanced equipment allows individuals to explore new possibilities and push the boundaries of creativity, resulting in more innovative and captivating content.*"

However, he asserted that having access to advanced equipment and software is insufficient; as individuals need proper training to use these tools effectively. Atiol asserted, "*Providing ample time for training is essential to ensure that individuals can use the software and equipment to their full potential.*"

Atiol also voiced his criticism regarding the limited time frame allocated for completing the final project in RNTC training. According to him, the limited time given for planning, scripting, shooting, and editing does not facilitate the production of high-quality content. Allowing more time for these crucial stages will contribute to the overall quality and success of the participants' projects.

Providing Ongoing Support

Participants in the training program emphasized the significance of offering continuous support to content creators even after the program concludes. This support can take various forms, including access to relevant resources, online platforms for discussing challenges and sharing best practices, and opportunities for further training and mentorship. Elias contributed, stating, "*By providing ongoing support, content creators can continually enhance their skills and knowledge, particularly in addressing and combating online hate speech.*"

Using RNTC as an example, it is fortunate to have a vast network of over 6,000 alumni from different parts of the world who can offer such

support. The organization has received positive feedback from its alumni about how their training at RNTC has propelled their careers and improved their work. Many alumni have achieved notable professional growth after completing a course at RNTC, with some even receiving awards. Therefore, RNTC alumni can serve as a valuable resource for content creators seeking ongoing support in their efforts to counter online hate speech.

Discussion

This study concluded that training content creators through critical digital training programs is an effective method for countering online hate speech, which supports the results of previous research by Gifford [34] that found that equipping individuals who create content with the necessary knowledge and skills to identify and address hate speech can promote respectful online environments, and combat the dangerous ideologies.

Participants in the present study also proposed that educating social media users and encouraging them to develop critical thinking skills are considered primary defense in the face of online hate. This notion aligns with the findings of a prior study by Foss SK, et al. [28], which found that people with strong critical digital literacy skills were more capable of identifying and reporting instances of hate speech on the internet.

The study underscored the significance of compassion and validation as key strategies in addressing online hate speech. Furthermore, the emphasis on finding common ground to foster communication and respect is consistent with findings from Cornwall's research on the effectiveness of empathy-based counter-speech messages [31]. Both studies highlight the potential of a humanized approach in cultivating a safe environment for open communication and mutual understanding.

Customizing counter/alternative narratives to suit the target audience's unique characteristics and cultural context were also one of the results that aligned with the results of past literature, specifically a study by Haider [30]. However, the current study highlighted the effectiveness of emotional appeals in increasing involvement with counter-messaging efforts. Moreover, it favored the creation of new peaceful

and inclusive narratives instead of countering existing ones and pointed out the advantages of doing so.

According to existing research, misinformation and harmful conspiracy theories tend to spread when hate speech is prevalent [11]. Meanwhile, the respondents in this study, highlighted the significance of being cautious about the sources of information and ensuring its accuracy through fact-checking. They believed that misinformation, disinformation, and mal-information can all be utilized to promote stereotypes and spread hateful rhetoric against certain individuals or groups, particularly minorities. This, in turn, can fuel the propagation of online hate speech.

In line with the findings of another previous study [28], where respondents recognized a broader connection between online and offline social realms, participants of this study expressed that online and offline worlds are intertwined. They expressed that a society marred by racism, tribalism, or polarization inevitably influences the behavior of its individuals online. Furthermore, they highlighted the concerning possibility that online hate can manifest itself in offline hate incidents.

To put it another way, individuals now recognize that their actions online can have real-world consequences, and that the virtual world is an extension of the physical world. They have come to understand that the internet is not a separate entity, but rather a reflection of society's values, attitudes, and behaviors. Therefore, it is important to consider the impact of online behavior on offline society and vice versa. This awareness can aid in combating hate speech and other forms of online harassment, as well as promoting a more tolerant and inclusive society both online and offline.

Conclusion

This study aimed to gain a deeper understanding of the experiences and perspectives of content creators who participated in the Radio Netherlands Training Centre (RNTC) training program entitled 'Digital Media to counter disinformation and hate speech' that took place in September/October 2022. Through the analysis of in-depth semi-structured interviews, the study addressed six research. These questions

explored participants' perceptions and evaluations of the program's effectiveness, their application of acquired knowledge and skills, effective learning techniques implemented by the trainers, the diffusion of the training materials to fellow content creators, recommended procedures for countering hate speech and creating a safe and inclusive online environment, in addition to their recommendations for improving digital literacy training programs to better combat online hate speech.

The results showed that the audience mapping and storytelling techniques taught in the training course were valued by all participants as the most beneficial aspects. The majority of the participants reported that they developed proficiency in recognizing hate speech, assessing its effects, and crafting counter-arguments and alternative arguments to combat it. Additionally, they mentioned that they acquired considerable knowledge on some social media moderation tools, most importantly social listening, in addition to thorough understanding of amplification and virality.

Participants recognized group exercises as the most effective learning technique employed by RNTC trainers, followed by presentations, open discussions, and demonstrations. The study further revealed that all participants successfully applied the knowledge and skills obtained from the RNTC training course. Their interventions against hate speech primarily involved raising awareness about its consequences, monitoring media, responding to instances of hate speech, and creating content that promotes peace and inclusivity. The study also shed light on the impact of hate speech on religious minorities and women, underscoring how it perpetuates false narratives and beliefs, fosters animosity, and contributes to the formation of stereotypes, prejudice, and discrimination.

Regarding the diffusion of the training materials to other content creators, results showed that 60% of participants shared the materials with their colleagues, students, professors, and junior trainees. This indicates the effectiveness of the training program in providing valuable and practical knowledge and skills that participants deemed worthy to pass on to others. It also suggests that the diffusion of training materials

can have a wider impact, as more individuals and organizations become equipped with the necessary tools to create high-quality content and combat issues like online hate speech.

Participants recommended best procedures for content creators to counter online hate speech. These procedures included promoting critical thinking skills and educating social media users about hate speech, activating fact-checking and verification tools to prevent the spread of biased, misleading or harmful content, creating tailored counter/alternative narratives for diverse target audiences, ensuring better representation and participation for marginalized communities online, reporting hateful content to social media platforms, harnessing the Power of AI to combat hate speech, and keeping an open-mind to avoid meaningless feuds.

Additionally, the study elicited six key recommendations from participants to enhance media and digital literacy training programs aimed at helping content creators counter online hate speech. These recommendations encompass providing trainees with a strong theoretical foundation, recognizing the diversity within developing countries rather than treating them as a homogeneous group, considering the participants' cultural context during the training process, aligning the training content with the needs and expectations of trainees, improving digital literacy programs through advanced facilities, equipment, and sufficient training time, and offering ongoing support to content creators even after the training program concludes.

Recommendations for future research

Based on the findings of the current study, the researcher suggests some potential avenues for future research, outlined below:

Long-term effectiveness: Conduct a follow-up study to assess the long-term effectiveness of the RNTC training program in countering disinformation and hate speech. This can involve tracking participants' progress in applying the knowledge and skills acquired from the training program, as well as evaluating the impact of the training on their career trajectories.

Comparative analysis: Compare the effectiveness of the RNTC training program with other similar training programs offered by different organizations in different countries. This can involve assessing the similarities and differences in the training content, delivery, and outcomes, as well as identifying best practices for designing and implementing effective training programs.

Participant diversity: Explore the experiences and perspectives of participants from diverse backgrounds, such as individuals from marginalized communities, religious minorities, or those with disabilities. This can involve assessing the effectiveness of the training program in addressing the specific challenges and issues faced by these groups, as well as identifying strategies to make the training program more inclusive and accessible.

Digital literacy and media education: Investigate the broader issues of digital literacy and media education in the context of countering disinformation and hate speech. This can involve exploring the role of schools and universities in promoting digital literacy and media education, as well as identifying best practices for integrating these skills into the curriculum.

Social media platforms: Examine the role of social media platforms in promoting or countering disinformation and hate speech. This can involve assessing the effectiveness of existing moderation policies and tools, as well as identifying strategies to improve the platform's ability to counter disinformation and hate speech.

AI and machine learning: Investigate the use of AI and machine learning techniques in countering disinformation and hate speech. This can involve exploring the potential of these technologies to automate the detection and removal of harmful content, as well as identifying ethical and legal issues associated with their use.

References

1. Perrin, A. Social media usage: 2005-2015. Pew Research Center; 2016.

2. Chirwa, E. Addressing online hate speech in developing countries: challenges and opportunities. Int J Hum Rights. 2018; 22(2):261-277.

3. Lichterman, J. The role of civil society in countering online hate speech in developing countries. J Hum Rights Technol. 2019; 1(2):143-163.

4. Brouwer, M., Kigongo-Bukenya, J., & Uwineza, J. Online hate speech in Africa: trends and challenges. Inf Technol Int Dev. 2020; 16(2):121-135.

5. Gangadharan S, Tan C. Internet censorship in developing countries: A comparative study of India and China. Telematics and Informatics. 2018 Feb 1; 35(1):9-19.

6. MacAvaney S, Yao H-R, Yang E, Russell K, Goharian N, Frieder O. Hate speech detection: Challenges and solutions. PLoS ONE. 2019 Aug 1; 14(8):e0221152. https://doi.org/10.1371/journal.pone.0221152.

7. Nockleby JT. Hate Speech. In: Encyclopedia of the American Constitution. 3rd Ed. New York: Macmillan Reference USA; 2000. p. 1277-9.

8. Brink A, Nelissen RAM. The dark side of online interaction: Anonymity, temptation, and the spreading of hate speech. Journal of Applied Social Psychology. 2019 Apr; 49(4):229-236.

9. Hatebase. Hate speech statistics. http://www.hatebase.org/popular. Accessed [12/2/2023].

10. Jackson TD, Nafus D. Online hate speech and its effects: A review of the literature. Comput Human Behav. 2020;110:106613.

11. Phillips W, Milnes T. The regulation of hate speech online: A comparative analysis of the United Kingdom, the United States, and Germany. Int J Law Inf Technol. 2017;25(3):238-257.

12. Peace makers network. EQUIPPING PAKISTANI YOUTH TO COUNTER ONLINE HATE SPEECH DURING THE COVID-19 CRISIS: A CASE REVIEW. November 2021. https://www.peacemakersnetwork.org/wp-content/uploads/2022/04/equipping-pakistani-youth-to-counter-hate-speech.pdf. Accessed [12/2/2023].

13. Hidosi A, Khadija B. Hate speech through social networking sites and its repercussions on Algerian youth - a field study. Al Resala Journal for Human Studies and Research. 2018;7(2):232-248. https://www.asjp.cerist.dz/en/article/184771. Accessed [10/2/2023].

14. Shaker F. Hate speech and racism in social media and methods of prevention from the point of view of students of the University of M'sila A field study on students of the departments of psychology and sociology [master's thesis]. The University of M'sila; 2020. http://dspace.univ-msila.dz:8080//xmlui/handle/123456789/24693. Accessed [7/2/2023].

15. Khasawna AS. The role of social media in spreading hate speech from the viewpoint of Jordanian journalists: a field study. J Islam Univ Sharia Legal Stud. 2021;29(1):300-322. ISSN 2148-2616.

16. Abdel Razek H. Determinants of Egyptian youth awareness of the concept of hate speech and its strategies in digital media. Arab J Media Commun Res. 2020 Apr-Jun;(29). Available from: https://jkom.journals.ekb.eg/article_106972.html?lang=en.

17. Timberg, C., & Chiu, A. (2020, April 8). As the coronavirus spreads, so does online racism targeting Asians, new research shows. The Washington Post. Retrieved from https://www.washingtonpost.com/technology/2020/04/08/coronavirus-spreads-so-does-online-racism-targeting-asians-new-research-shows/

18. Fladmoe, A., & Nadim, M. (2019). Silencing Women? Gender and Online Harassment. Social Science Computer Review, 39(2), 1-14. https://doi.org/10.1177/0894439319865518

19. Vogels, E. A. (2021, January 13). The state of online harassment. Pew Research Center. Retrieved from https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/

20. Piñeiro-Otero, T., & Martínez-Rolán, X. (2021). Say it to my face: Analysing hate speech against women on Twitter. Profesional de la Información, 30(5), e300502. https://doi.org/10.3145/epi.2021.sep.02

21. Post, R., Hare, I., & Weinstein, J. (2009). Hate speech. In Extreme speech and democracy (pp. 123-138). Oxford University Press.

22. HRC. (2015). Report of the Special Rapporteur on minority issues, Rita Izsák. A/HRC/28/64.

23. United Nations. (2019). United Nations Strategy and Plan of Action on Hate Speech. Retrieved from https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf

24. Scitt, M. Social media companies remove less hate speech in 2021. POLITICO. October 7, 2021. Retrieved from: https://www.politico.eu/article/facebook-google-hate-speech-social-media-european-commission-transparency/

25. Frenkel S, Conger K. Hate Speech's Rise on Twitter Is Unprecedented, Researchers Find. New York Times. Dec. 2, 2022. Available from: https://www.nytimes.com/2022/12/02/technology/twitter-hate-speech.html

26. Mouratoglou N. Combating Online Hate Speech through Critical Digital Literacy: Reflections from an Emancipatory Action Research with Roma Youths. Journal of Media Literacy Education [Internet]. 2021 [cited 2023 Mar 12];11(2). Available from: https://digitalcommons.uri.edu/jmle/vol11/iss2/7/

27. Hinduja S, Patchin JW. Bullying, cyberbullying, and suicide. Archives of Suicide Research. 2010;14(3):206-221. doi: 10.1080/13811118.2010.494133.

28. Foss SK, et al. The Role of Critical Digital Literacy in Recognizing and Reporting Hate Speech. Journal of Computer-Mediated Communication. 2020;25(4):1-19.

29. Zeiger S, Gyte J. Prevention of Radicalization on Social Media and the Internet. In: Rubin GJ, Wessely S, editors. Handbook of Terrorism Prevention and Preparedness. International Centre for Counter-Terrorism (ICCT); 2020. Chapter 12.

30. Haider H. Countering online misinformation, hate speech or extremist narratives in the Global South. K4D Helpdesk Report. 2022. Available from: https://opendocs.ids.ac.uk/opendocs/bitstream/handle/20.500.12413/17464/1113_Countering_online_misinformation_hate_speech_extremism.pdf?sequence=1. Accessed February 20, 2023.

31. Cornwall I. Combating Hate Speech Through Media: A Young Caribbean Woman's Perspective. United Nations. 2022. Available from: https://www.un.org/en/academic-impact/countering-hate-speech-through-media-young-caribbean-woman%E2%80%99s-perspective. Accessed February 20, 2023.

32. Yzer MC, Southwell BG, Kim J. Empathy-based and other strategies to reduce xenophobic hate speech on Twitter: A randomized controlled trial. PLoS One. 2021;16(1):e0244237. doi: 10.1371/journal.pone.0244237.

33. IDEAS (Inclusive Development Alternative Solutions) by Search for common Ground–Sri Lank. Cyber Guardians: Empowering Youth to Combat Online Hate Speech in Sri Lanka. June 2019–September 2019.

34. Gifford, R. Environmental psychology principles and practice. 5th ed. John Wiley & Sons; 2016.

35. Lin, C.-Y., Hasson, U. The impact of training on reducing hate speech in online social media. Cyberpsychol Behav Soc Netw. 2019;22(9):554-560. doi: 10.1089/cyber.2019.0101.

36. Kim, J., et al. Combating Hate Speech with Machine Learning: An Analysis of the Effectiveness of Automatic Hate Speech Detection Systems. J Comput Sci Technol. 2018;12(4):1-12.

37. Chaudhary, M., Saxena, C., Meng, H. Countering Online Hate Speech: An NLP Perspective. The Chinese University of Hong Kong; 2021. Available from: https://www.researchgate.net/publication/354434964_Countering_Online_Hate_Speech_An_NLP_Perspective#pf. Accessed February 2, 2023.

38. Melki J. Journalism and Media Studies in Lebanon. Journalism Studies. 2009;10(5):672-690. doi: 10.1080/14616700902920174.

39. Home. [Digital Skills Camp] Facebook. Available from: https://www.facebook.com/profile.php?id=100070514851883. Accessed February 2, 2023.

40. RNTC. Available from: https://rntc.com/. Accessed March 1, 2023.

41. Melki J, Maaliki L. Helping Arab Digital and Media Literacy Blossom. Three years of the Media and Digital Literacy Academy of Beirut (MDLAB).

Opportunities for Media and Information Literacy in the Middle East and North Africa. Yearbook 2016. doi:10.13140/RG.2.2.22529.81764.

42. Levitt HM, Motulsky SL, Wertz FJ, Morrow SL, Ponterotto JG. Recommendations for Designing and Reviewing Qualitative Research in Psychology: Promoting Methodological Integrity. Qualitative Psychology. 2017 Mar;4(1):2-22.

43. Zohrabi M. Mixed Method Research: Instruments, Validity, Reliability and Reporting Findings. Theory and Practice in Language Studies. 2013 Feb;3(2):254-62.

44. Tong A, Flemming K, McInnes EF, Craig J. Enhancing Transparency in Reporting the Synthesis of Qualitative Research: ENTREQ. BMC Med Res Methodol. 2012 Nov;12:181.

45. Rebolj AB. The case study as a type of qualitative research. Journal of contemporary educational studies. 2013;1. Available from: https://www.researchgate.net/publication/265682891.

46. Kvale S, Brinkmann S. Interviews: Learning the Craft of Qualitative Research Interviewing. 2nd ed. Los Angeles, CA: Sage Publications; 2009.