Combating Online Violent Extremism Through AI: Avenues for Pakistan

Nimra Javed[*]

Abstract

This research paper explores the potential utilization of Artificial Intelligence (AI) as a tool to counteract online violent extremism. The study specifically concentrates on the Pakistani landscape, recognizing the imperative for proactive measures due to the constantly evolving nature of violent extremism, particularly in the realm of internet and social media. It uses qualitative methodology as it relies primarily on extensive literature review and logical conclusions drawn from relevant research in the field. The primary focus of this inquiry lies in scrutinizing the dual-use predicament associated with AI, wherein technology serves both advantageous and detrimental roles. Additionally, the research delves into the strategies employed by extremist entities in exploiting online channels for recruitment and radicalization. The core purpose of this study however revolves around utilization of AI for effective identification and mitigation of extremist content. This objective is pursued through the application of sentiment analysis and threat intelligence mechanisms, which monitor activities across the dark web and offer anticipatory insights into potential threats. A crucial proposition set forth by this study involves the development of AI detection models tailored to the nuances of the Pakistani context.

Keywords: *Artificial Intelligence, Machine Learning, Violent Extremism, Social Media, Sentiment Analysis, Terrorism.*

[*] Nimra Javed is an Associate Research Officer at Center for International Strategic Studies AJK. She has done M.Phil. in Strategic Studies from National Defence University, Islamabad and a has a number of national and international publications. Her area of interest is emerging technologies & new trends in warfare period. She can be reached at nimrahjaved42@gmail.com

1. Introduction

The pervasiveness of social media coupled with its easy and hassle-free access has provided extremists in Pakistan and the world over with a handy tool by means of which they can promote and advertise their highly controversial and dangerous content. Many extremist organizations have quite skillfully leveraged social networking sites and video-sharing platforms such as Facebook, Twitter, and YouTube to disseminate their ideologies and allure prospective adherents. Owing to the user-friendly nature of these platforms, there has been an expeditious proliferation of extremist content. This content encompasses a fairly wide spectrum, ranging from expressions of hatred and fervent narratives to extremely graphic depictions of violent perpetrations.[1] This online proliferation has markedly exacerbated the predicament confronted by Pakistani authorities in mitigating the surge of violent extremism.

Violent Extremists can be defined as, "individuals who support or commit ideologically motivated violence to further their political goals."[2] Violent Extremism (VE) is multifaceted as it may have various manifestations like religious extremism, anti-government, right-wing and left-wing, for which social media platforms can be conveniently used. This necessitates countering such trends through the use of technological innovations such as Artificial Intelligence (AI). AI, here is understood as the narrow AI and not Artificial General Intelligence (AGI).[3]

Narrow AI systems are tailored for specific tasks and lack the comprehensive cognitive abilities typically sought in AGI. In comparison, broad

[1] Marvin G. Weinbaum, "Insurgency and Violent Extremism in Pakistan," *Small Wars & Insurgencies* 28, no. 1 (2017): 34–56. https://doi.org/10.1080/09592318.2016.1266130

[2] White House, *Empowering Local Partners to Prevent Violent Extremism in the United States* (Washington, D.C.: Home Land Security Report, 2011), 1.

[3] Scott M. Martin Kane James R. Casey, Stephanie, *History of Artificial Intelligence and Personalized Learning* (United Kingdom: Routledge, 2021).

AGI systems aspire to encompass a wide array of capabilities. Narrow AI systems are engineered with the purpose of executing particular functions and often necessitate substantial reprogramming or retraining to undertake alternative tasks. The implementation of AI technology carries the potential for both positive and adverse consequences. Terrorist organizations may harness AI to amplify the dissemination of their propaganda. Conversely, government entities entrusted with law enforcing responsibilities, may leverage AI to enhance their operational efficacy and counter extremist strategies.[4]

To effectively combat the scourge of violent extremism, law enforcement agencies are progressively adopting AI technology. AI algorithms possess the capability to discern patterns and indicators associated with extremist behavior through the analysis of extensive volumes of online data, encompassing content from social media platforms, chat records, and websites, among other digital sources.

Machine learning (ML) based algorithms facilitate swift identification and surveillance of potentially incriminating information, thereby empowering authorities to proactively intervene and mitigate potential threats. Furthermore, the incorporation of AI-driven sentiment analysis and linguistic processing aids in deciphering the tenor and intent of online communications, thereby aiding law enforcement agents in distinguishing innocuous discourse from potentially perilous expressions. This symbiotic amalgamation of AI and law enforcement capacities augments the ability to detect and counteract instances of violent extremism, thereby enhancing the safety of both the digital realm and the society at large.[5] Thomas James Vaughan Williams et al., argue that social media plays an extremely

[4] MaryAnne M. Gobble, "The Road to Artificial General Intelligence," *Research-Technology Management* 62, no. 3 (2019): 55–59, https://doi.org/10.1080/08956308.2019.1587336.

[5] David W. Bates et al., "Reporting and Implementing Interventions Involving Machine Learning and Artificial Intelligence," *Annals of Internal Medicine* 172, no. 11 Suppl (2020): 37–44, doi: 10.7326/M19-0872.

important role in modern-day communications.[6] Hanna Munden in her article similarly discusses the role of social media in transforming people's lives and even identities.[7]

This study examines the spread of violent extremism in Pakistan on social media platforms as well as discusses how AI can be used to counter violent extremism in online spaces. It is important to note at the start that AI can be used for both spreading and combating extremism online. It is crucial therefore to invest resources into not just how AI can be utilized for effectively countering violent extremism online but also how terrorist and extremist organizations can be barred from using AI for their nefarious intentions.

2. Methodology

This study takes a qualitative research methodology, utilizing secondary data sourced from scholarly publications, government documents, and online archives. The methods employed for data collection encompass a comprehensive literature review and thematic analysis, placing a significant emphasis on the identification and in-depth understanding of overarching themes pertinent to the study.

3. Dual Use Dilemma

Historically, the term "dual-use" has been employed to delineate technologies with the capacity to serve both civilian and military enterprises. In the contemporary technological landscape, the conundrum of dual use arises when a scientific pursuit has benevolent as well as malevolent applications. This predicament, often referred to as the "dual-use dilemma," encapsulates instances where a technological artifact,

[6] Thomas James Vaughan Williams et al., "Policy vs Reality: Comparing the Policies of Social Media Sites and Users' Experiences, in the Context of Exposure to Extremist Content," *Behavioral Sciences of Terrorism and Political Aggression* (2023): 1–18, https://doi.org/10.1080/19434472.2023.2195466.

[7] Hanna Paalgard Munden, "Extremist Group Exits: What Autobiographies by Male Right-Wing Formers Reveal about Identity Transformation," *Behavioral Sciences of Terrorism and Political Aggression* (2023): 1–22, https://doi.org/10.1080/19434472.2023.2192771.

tool, or body of knowledge could be harnessed to yield positive or negative outcomes, contingent upon its application.[8]

The nomenclature "dual-use dilemma" emerges from the broader concept of "dual-use," representing a scenario characterized by the intricate interplay of two opposing applications or implications. This intricate paradox manifests when an innovation, conceived with laudable objectives in mind, inadvertently presents the possibility of being exploited or repurposed to perpetrate acts that are inherently deleterious or injurious.[9]

Numerous technologies embody the potential for dual use, wielding the capacity to ameliorate human existence, advance scientific cognition, and propel economic advancement. Often borne from altruistic motives, these technologies find application in domains such as medical advancement, environmental monitoring, communication enhancement, and mitigation of the afflictions stemming from natural calamities.[10]

These very technologies may be susceptible to malevolent exploitation, encompassing realms like surveillance, cyber warfare, biological weaponry, and the augmentation of numerous military capabilities, among others. The potential for detriment may arise from unforeseen consequences, malevolent manipulation of technology, or the perpetual evolution of technical prowess.[11]

[8] Amir Lupovici, "The Dual-Use Security Dilemma and the Social Construction of Insecurity," *Contemporary Security Policy* 42, no. 3 (2021): 57–85, https://doi.org/10.1080/13523260.2020.1866845.

[9] James Giordano, ed., *Neurotechnology in National Security and Defense: Practical Considerations, Neuroethical Concerns* (United States: CRC Press, 2014).

[10] Guangyou Zhou et al., "Inclusive Finance, Human Capital and Regional Economic Growth in China," *Sustainability* 10, no. 4 (2018): 1194. https://doi.org/10.3390/su10041194

[11] Marcus Comiter, "Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It," Belfer Center for Science and International Affairs, last updated August 2019, https://www.belfercenter.org/publication/AttackingAI.

The ethical quandary of selecting between two divergent applications invokes moral apprehensions concerning obligation, culpability, and intent. Innovators and researchers, perhaps unwittingly, become enmeshed in endeavors yielding adverse ramifications. Deliberate ethical ratiocination becomes imperative to strike an equilibrium between the potential for positive impact and the potential for negative repercussions. The ramifications of dual-use technologies extend beyond individual ethical predicaments, imperiling global stability and security. The utilization of AI presents a dichotomy of beneficial and detrimental applications. This paradox is particularly evident when considering how an algorithm with substantial economic advantages could inadvertently lead to the creation of massively destructive weaponry. Globally, AI's misuse harbors the potential for catastrophic consequences.

4. Spread of Violent Extremism through Internet

Researchers are widely in consensus regarding the immense potential of online indoctrination. Substantial evidence indicates that violent extremist organizations exploit online channels for multifarious activities, encompassing recruitment, propagation of propaganda, community cultivation, psychological warfare, tactical planning, information dissemination, network establishment, and financial transactions.[12] The modalities employed for these activities' manifest variability contingent on the group's ideology and individualistic considerations.

Extremists advocating violence are capitalizing on the facile accessibility to an ever-expanding audience facilitated by the internet, thereby recruiting, grooming, and radicalizing vulnerable individuals. Social media in particular

[12] Ryan Scrivens et al., "Comparing the Online Posting Behaviors of Violent and Non-Violent Right-Wing Extremists," *Terrorism and Political Violence* 35, no. 1 (2023): 192. https://doi.org/10.1080/1057610X.2022.2099269.

allows extremist recruiters to exponentially broaden their outreach to groups and individuals hitherto unreceptive to conventional methodologies.[13]

This engenders the potential for radical recruiters to engage with otherwise inaccessible segments of the population. Their strategies involve promulgating their ideologies, fomenting antipathy towards adversaries, exhorting acts of violence, venerating martyrs, constituting virtual enclaves of like-minded adherents, furnishing religious or legal rationale for proposed actions, and interfacing with and indoctrinating novices. These activities are carried out on a range of online platforms, encompassing ordinary websites, mainstream social media conglomerates such as Facebook, Twitter, and YouTube, alongside other digital services.[14]

Dissemination of incendiary content, including instructional materials elucidating bomb fabrication and weaponry operation, recordings chronicling successful assaults, discourses espousing radical doctrines, blog entries, and comments endorsing and further inflaming acts of aggression constitute prevalent strategies espoused by online extremists. Notably, terrorist entities exploit the popular online social media platform, Facebook, for confidential communication and information exchange to coordinate attacks, as well as generating pages garnering user "likes" to showcase support. Furthermore, these entities harness the likes of Twitter for disseminating propaganda and official statements, alongside uploading radical sermons and instructional videos onto YouTube.

A notable illustration pertains to the Islamic State of Iraq and al-Sham (ISIS), recognized for its adept production of sophisticated audiovisual content, extensively propagated across prominent social media avenues such as YouTube

[13] Paul Cornish, ed., *The Oxford Handbook of Cyber Security* (England: Oxford University Press, 2021).

[14] Patricia R. Recupero and Samara E. Rainey, *The Internet and Social Media as an Enabling Force* (England: Oxford University Press, 2022).

and Twitter. Furthermore, ISIS has adeptly assimilated emergent technologies and embraced evolving social media platforms, notably including Telegram. These endeavors collectively serve the purpose of amplifying its ideological discourse and engendering the recruitment of fresh adherents within the realm of cyberspace.[15]

Their interactions, networking, and message dissemination have transpired via diverse internet mediums, comprising message boards, chat rooms, and even dating websites. By establishing profiles, pages, and accounts across a spectrum of platforms, smaller organizations project an illusion of critical mass and devoted participation for their cause. Furthermore, the pervasive availability of internet connectivity across global regions amplifies the appeal of these organizations and project an impression of heightened potency.

5. Use of AI by Extremist Groups

Extremist groups may use AI for various notorious activities. Deepfakes, being one of the most stand out of such activities, constitute fabricated audio and/or visual content that has been manipulated or synthesized through the utilization of Generative Adversarial Networks (GANs). The term "deepfakes" pertains to this genre of manipulated media. Deepfakes have emerged as a prominent exemplar of the untoward employment of AI in contemporary times, garnering significant attention from media due to the conundrums they pose in distinguishing authentic renditions from counterfeit ones (a challenge encountered by both human observers and computational systems). This technology and its associated potential have cast a shadow over the landscape of information warfare, wherein deepfakes could serve as potent instruments.

[15] Yasmin Ibrahim, *The Sharing Economy and Livestreaming of Terror: Co-Production of Terrorism on Social Media* (England: Oxford University Press, 2021).

The credibility of conventional, authoritative media is poised to undergo erosion in the wake of the pervasive use of deepfakes for dissemination of misleading information. The intricate task of verifying the authenticity of videos creates a situation wherein any conceivably compromising visual material can be debunked, given the plausible assertion that the audio-visual content has been synthetically fabricated. Consequently, even if the fidelity of the contested video material remains unscathed, possibility for any party to repudiate its genuineness persists, thereby enabling them to evade accountability for the objectionable actions.

Beyond the realm of audio-visual deepfake creation, AI can be harnessed for generation of tailored narratives with the intent of fomenting radicalization. Apprehensions have been raised regarding the potential exploitation of this technology for micro-profiling, micro-targeting, automatic composition of persuasive text for recruitment endeavors, and dissemination of customized counterfeit news and conspiracy theories linked to terrorism. The advent of sophisticated techniques in Natural Language Processing (NLP), notably exemplified by Open AI's prominently featured GPT-3, has further exacerbated these concerns.

The utilization of AI-fueled spurious news platforms may yield deleterious consequences, particularly considering the escalating proclivity among online readers to disseminate articles based on their headlines, often engaging in cursory perusal without undertaking substantive due diligence regarding the veracity of the sources. This trend is especially disconcerting, given its growing prevalence. Consequently, the prospect arises that extremist organizations might eventually deploy artificial intelligence systems that autonomously analyze genuine news headlines and generate succinct yet fabricated proclamations for propagation across social media and analogous conduits, furthering their ideological agenda.

6. Use of AI to Counter Violent Extremism

Automated content moderation has gained popularity as a pragmatic approach to managing the overwhelming volumes of user-generated content online and the rapidity with which specific content can attain viral status. This strategy has proven effective in addressing the extensive amounts of user-generated content online. In the realm of content curation and moderation, private organizations employ diverse automated methods that involve removal, downgrading, or redirection of viewers to alternative content.[16] For example, Facebook utilizes ML algorithms to ascertain the priority level of each piece of content necessitating examination.[17]

Instances of content that contravene the company's regulations can be flagged by users or detected by ML filters. These violations encompass a spectrum from spam to content that "glorifies violence," including expressions of hate speech. Since 2020, Facebook has chosen to expeditiously handle blatant transgressions by promptly removing the content or suspending the associated account. Human content moderators are only tasked with reviewing content when reasonable suspicion arises regarding its alignment with the company's policies.[18]

In its campaign against terrorism and violent extremism, Facebook employs a spectrum of strategies, among which is the utilization of AI language models to comprehend content that may endorse terrorism. Such textual content is often specific to distinct languages and socio-cultural groups.[19]

[16] Nathaniel Persily, *Platform Power, Online Speech, and the Search for New Constitutional Categories* (England:Oxford University Press, 2022).

[17] PRC, "Code-Dependent: Pros and Cons of the Algorithm Age," (*Pew Research Center: Internet, Science & Tech,* 2017), https://www.pewresearch.org/internet/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/.

[18] Thomas Stackpole, "Content Moderation Is Terrible by Design," *Harvard Business Review,* 2022. https://hbr.org/2022/11/content-moderation-is-terrible-by-design.

[19] Brian Fishman, "Dual-Use Regulation: Managing Hate and Terrorism Online before and after Section 230 Reform," *Brookings,* March 14, 2023, https://www.brookings.edu/articles/dual-use-regulation-managing-hate-and-terrorism-online-before-and-after-section-230-reform/.

Notwithstanding its drawbacks, automated content moderation systems are increasingly acknowledged as an imperative within the private sector. This is owing to the prodigious volume of information disseminated online daily. Given the inherent intricacies, it is unsurprising that governmental authorities traditionally possess limited influence over content moderation and removal.[20] In response to mounting criticisms that social media corporations inadequately manage information proliferation within prevailing self-regulatory frameworks, several nations have endeavored to enact national legislation aimed at compelling companies to take more decisive actions.[21]

This could encompass regulating the pace at which content must be expunged alongside introducing incentives and penalties for violations. A notable instance of such national legislation is Germany's Network Enforcement Act, 2017. Another illustrative example is the European Union's recent legislative measures to combat the dissemination of terrorist content online, which mandates the removal of such content within an hour of receiving removal requests.[22]

7. Sentiment Analysis

Individuals utilize online social networking platforms to openly express their thoughts, viewpoints, and emotions, whether positive or negative. Prior research has examined the sentiments conveyed on these platforms to investigate behaviors across diverse settings and for various purposes. Researchers have been

[20] Ozlem Ozmen Garibay et al., “Six Human-Centered Artificial Intelligence Grand Challenges,” *International Journal of Human–Computer Interaction* 39, no. 3 (2023): 391–437. https://doi.org/10.1080/10447318.2022.2153320.

[21] Chang Sup Park and Homero Gil De Zúñiga, “Learning about Politics from Mass Media and Social Media: Moderating Roles of Press Freedom and Public Service Broadcasting in 11 Countries,” *International Journal of Public Opinion Research* 33, no. 2 (2021): 15–35. https://doi.org/10.1093/ijpor/edaa021.

[22] Library of Congress, “Germany: Network Enforcement Act Amended to Better Fight Online Hate Speech,” Law Library of Congress, Washington, D.C., July 6, 2021, https://www.loc.gov/item/global-legal-monitor/2021-07-06/germany-network-enforcement-act-amended-to-better-fight-online-hate-speech/.

particularly interested in employing automated techniques to classify the polarity of public sentiments based on concise language usage in communications, such as tweets. This involves the analysis of data gathered from social media platforms to glean insights into public opinion.

Facebook, Twitter, Tumblr, and YouTube exemplify online social networks that have evolved into prominent platforms for millions of individuals utilizing the internet to establish and nurture interpersonal connections. The proliferation of microblogging services in recent years has significantly influenced people's modes of thinking, communicating, behaving, learning, and conducting businesses. These widely adopted social platforms represent innovative forms of blogging developed to facilitate enhanced interpersonal communication. Users express their beliefs, ideas, and thoughts, constructively or detrimentally, whether through tweet messages, blog posts, article and video shares, or messages on social networking sites.

In their 2019 study, Ahmad et al. employed a binary classification method to detect potential extremist affiliations. The focal point of this research is machine learning classifiers, encompassing models like random forests, support vector machines (SVMs), k-nearest neighbors (KNNs), Naive Bayes, and deep learning. This investigation is structured into three primary segments, deploying emotion-based extremist categorization techniques to analyze tweets:

i. Collection of user tweets;

ii. Classification of these tweets into extremist and non-extremist groups using various deep learning-based sentiment models;

iii. Preparation and processing of the data.

These models integrate techniques such as FastText, Convolutional Neural Networks (CNN/ConvNet), and Long Short-Term Memory (LSTM) networks.[23]

While the study's results demonstrate the method's efficacy in enhancing precision, recall, F-measure, and accuracy, certain limitations persist- such as the classification is confined to a binary class categorization rather than a multi-class scheme. Furthermore, the process of gathering, cleaning, and storing Twitter data lacks the desired level of automation required to streamline these procedures and enhance efficiency.

In Kaur et al.'s study, deep learning is employed to automatically detect extremism. Annotators specializing in relevant themes were engaged to categorize collected data into three classes: radical, non-radical, and irrelevant. Word2Vec was utilized to generate word embeddings from the collected data. To identify extremism and classify data as radical, non-radical, or irrelevant, the study utilized LSTM. The data was labeled by specialized annotators based on attributes provided by the authors.[24] These attributes encompass aspects such as mistreatment of military personnel, anti-national rhetoric, promotion of terrorism or terrorists, and incitement of others. The identification of radical content across online media employed various ML techniques including random forest, SVM, and Max Entropy. The proposed method achieved an accuracy of 85.7%, which could be further refined by incorporating an additional layer of CNN to enhance attribute identification precision.[25]

[23] Shakeel Ahmad et al., "Detection and Classification of Social Media-Based Extremist Affiliations Using Sentiment Analysis Techniques," *Human-Centric Computing and Information Sciences* 9, no. 1 (2019): 24. https://doi.org/10.1186/s13673-019-0185-6.

[24] Armaan Kaur, Jaspal K. Saini, and Divya Bansal, "Detecting Radical Text over Online Media Using Deep Learning," *ArXiv*, July 30, 2019, https://arxiv.org/abs/1907.12368.

[25] Kaur, Saini, and Bansal, "Detecting Radical Text over Online Media," 2

8. Spread of Violent Extremism in Pakistan

In the context of Pakistan's terrorist milieu, groups like Tehreek-e-Taliban Pakistan (TTP) and Daesh-Khorasan (Daesh-K) have established substantial social media propaganda operations. Despite AI-driven checks and significant purging efforts by platforms like Facebook and Twitter, TTP and Daesh-K have maintained a modest presence on these networks. Furthermore, both factions are increasingly utilizing encrypted chat platforms such as WhatsApp, Rocket Chat, Hoop Messenger, and Telegram.[26]

To circumvent crackdowns, both TTP and Daesh-K have transitioned much of their propaganda to the dark web. However, the proliferation of social media platforms poses significant challenges in completely eradicating TTP and Daesh-K propaganda from their accounts. These groups demonstrate remarkable adaptability, continually adjusting to the evolving social media landscape. Their survival and propaganda dissemination rely on constantly seeking new, lesser-known, yet secure encrypted platforms.[27]

While past studies have concentrated on detecting radical content or identifying social media accounts displaying radical indicators, it is paramount to underscore here the complexity of simply categorizing individuals as radical or extremist. The potential for false positives underscores the delicate nature of this task, as misclassification may subject innocents to unwarranted surveillance or policing scrutiny.

9. Violent Extremism and AI in Pakistan

The rise of AI and ML technologies holds both promising and challenging implications for Pakistan's efforts in countering violent extremism, particularly on

[26] Abdul Basit Khan, "Creating Social Firewalls against Militants' Online Propaganda Is Critical for Pakistan's Internal Security," *Arab News,* March 16, 2023, https://arab.news/vxv6p.

[27] Basit, "Creating Social Firewalls against Militants."

the dark web. Effective utilization of these technologies requires careful calibration, enabling cybersecurity and law enforcement professionals to adequately detect and mitigate threats. Although AI and ML excel in analyzing the extensive data on the dark web to identify patterns in terrorist activities and enhance real-time threat detection, their success heavily depends on a deep understanding of the local context- cultural and linguistic nuances. Therefore, a comprehensive strategy integrating technical expertise with human insight is crucial for Pakistan, acknowledging the limitations of existing AI models while maximizing their strengths to combat violent extremism effectively.

To safeguard citizens from the potential hazards posed by the dark web, law enforcement agencies must maintain a constant state of vigilance and harness evolving technologies such as AI and ML.[28] AI and ML offer a diverse array of sophisticated tools and strategies that law enforcement and cyber security experts can employ to more effectively monitor and counteract the risks emanating from the dark web. These technologies can effectively be leveraged to bolster security.[29]

For instance, these technologies can be adeptly utilized in the realm of threat intelligence. AI and ML can process extensive volumes of data sourced from the dark web, discerning intricate patterns and trends in criminal activities and terrorist organizations.[30] This extracted information can then serve as a foundation for guiding law enforcement operations and formulating cyber security measures that exhibit heightened efficacy. Moreover, real-time threat detection, powered by

[28] Randa Basheer and Bassel Alkhatib, “Threats from the Dark: A Review over Dark Web Investigation Research for Cyber Threat Intelligence,” *Journal of Computer Networks and Communications* 2021 (2021): 1–21. https://doi.org/10.1155/2021/1302999.

[29] Christopher Rigano, “Using Artificial Intelligence to Address Criminal Justice Needs,” Government Website, National Institute of Justice, October 8, 2018, https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs.

[30] Anand Singh Rajawat et al., “Dark Web Data Classification Using Neural Network,” *Computational Intelligence and Neuroscience* 2022 (2022), https://doi.org/10.1155/2022/8393318.

AI and ML, permits prompt identification of malicious behaviors such as malware proliferation and illicit trading of stolen data, terrorist activities and planning.[31] Consequently, rapid interventions can be undertaken to mitigate the perils stemming from these illicit activities.

An additional application lies in sentiment analysis, enabled by AI techniques. This entails scrutinizing the language used within dark web forums and other online communities to ascertain the prevailing tone and sentiment of discussions. This analysis can unveil potential threats, thereby furnishing law enforcement with insights that inform their strategic responses. In parallel, predictive analytics, a feature of AI and ML involves scrutinizing dark web data to anticipate forthcoming patterns and activities. By accessing this predictive capability, authorities can preemptively identify emerging threats and initiate preventive actions.[32]

In preparation for the deployment of a Threat Intelligence Platform (TIP), organizations must first delineate the specific categories of threats that necessitate monitoring. Thorough research is indispensable to select a fitting TIP that aligns with their objectives. Subsequently, a structured approach ensues, encompassing project scoping, TIP implementation, outcome assessment, action initiation, strategy review, and adaptive monitoring of the evolving threat landscape.[33] Dynamically adjusting TIP parameters to align with shifting threat contexts empowers businesses to remain ahead of potential dangers, affording them the

[31] Antonio João Gonçalves De Azambuja et al., "Artificial Intelligence-Based Cyber Security in the Context of Industry 4.0—A Survey," *Electronics* 12, no. 8 (2023): 1920. https://doi.org/10.3390/electronics12081920.

[32] Hsinchun Chen, "Sentiment and Affect Analysis of Dark Web Forums: Measuring Radicalization on the Internet," in *2008 IEEE International Conference on Intelligence and Security Informatics* (ISI 2008), Taipei, Taiwan: IEEE, 2008), 104–9. https://doi.org/10.1016/j.neucom.2015.09.063.

[33] Mario Faiella et al. "Enriching Threat Intelligence Platforms Capabilities." Paper presented in *16th International Proceedings of the Conference on Security and Cryptography (SECRYPT)* at Prague, Czech Republic, July 2019. doi: 10.5220/0007830400370048.

opportunity to refine their security architectures based on the discerned tactics, techniques, and procedures of threat actors.[34]

Efforts targeting the deep web take a similar methodical approach. Identifying pertinent data categories, conducting comprehensive research, selecting appropriate analysis tools, defining project prerequisites, tool configuration, result evaluation, and iterative refinement of tools contribute to an effective analysis process (in the context of AI deployment in Pakistan, it will become imperative to craft localized AI detection models).

TIP acknowledges that human expertise remains indispensable for a comprehensive understanding of the multifarious manifestations of terrorism and violent extremism across global contexts. Comparable prominent platforms, like Twitter and YouTube, similarly deploy AI mechanisms to swiftly eliminate comments that contravene their established guidelines.

The process of editing and refining content involves AI models being trained to disregard specific information based on predetermined criteria. Nevertheless, the current iterations of these models are encumbered by inherent limitations. For instance, an AI model trained to detect content from one terrorist group might prove ineffective for another due to linguistic and stylistic disparities in their propaganda. As highlighted by Karen Hao, an algorithm adept at identifying Holocaust denial might struggle to recognize, for instance, denial of the Rohingya genocide.[35] Education of artificial intelligence relies on data; its ability to filter content is contingent upon the availability of such data.

[34] Mario Faiella et al., "Enriching Threat Intelligence Platforms Capabilities,".

[35] Karen Hao, "How Facebook Got Addicted to Spreading Misinformation," *MIT Technology Review*, 2021.https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/.

AI also confronts substantial challenges when dealing with intricate language intricacies, notably sarcasm and humor, both of which can significantly impede the efficacy of automated content moderation. Presently accessible AI models are frequently tailored to major languages, rendering them less dependable for minority languages prevalent in regions such as South Asia and Southeast Asia.[36] In this particular context, complete automation of content moderation is either unfeasible or inadvisable. The process of information review and decision-making necessitates human oversight.

Furthermore, the nexus between limited employment prospects for educated individuals and the propensity for violent extremism merits attention. Recent analysis of foreign recruits by Daesh reinforces this viewpoint. The study, based on a 2016 report by Combating Terrorism Center (CTC) at West Point, examined data from approximately 4,000 Daesh foreign fighters, revealing above-average educational backgrounds juxtaposed with a prevalence of low-skilled employment histories.[37]

Therefore, the amalgamation of AI and ML technologies with vigilant law enforcement efforts presents a potent strategy for mitigating the threats of the dark web. The nuanced approaches described herein encapsulate the multifaceted nature of this endeavor, highlighting the need for tailored solutions that resonate within specific contexts.

It is imperative to consider the distinct cybersecurity challenges and socio-economic factors inherent to the region when assessing the utilization of AI and ML technologies in combating dark web threats in Pakistan. The proliferation of

[36] Greyson K. Young, "How Much Is Too Much: The Difficulties of Social Media Content Moderation," *Information & Communications Technology Law* 31, no. 1 (2022): 1–16. http://dx.doi.org/10.2139/ssrn.3792647.

[37] Arie Perliger and Daniel Milton, "From Cradle to Grave: The Lifecycle of Foreign Fighters in Iraq and Syria," *CTC at Westpoint* (2016). https://ctc.westpoint.edu/wp-content/uploads/2016/11/Cradle-to-Grave2.pdf

dark web extremists and cybercriminals poses unique risks to Pakistan's burgeoning digital infrastructure. Therefore, a primary focus for Pakistan's deployment of AI and ML should be the development of regionally tailored detection models proficient in regional languages and dialects. This is paramount for effective sentiment analysis and the acquisition of threat intelligence from South Asian online forums and communities.

Furthermore, it is noteworthy that a correlation exists between radicalization and the limited employment opportunities for college graduates. In this context, AI and ML can be instrumental in monitoring and tracking the online recruitment efforts of extremist organizations. To ensure the ethical application of these technologies and navigate the intricacies of linguistic and cultural diversity, it is imperative to establish robust regulatory frameworks and institute vigilant inspection.

The implementation of this customized approach holds the potential to enhance Pakistan's national cybersecurity and counter-terrorism operations significantly. This, in turn, will bolster the nation's capacity to proactively detect and neutralize dark web threats.

10. Conclusion

The proliferation of online violent extremism raises significant concerns across global societies, including Pakistan. Addressing these multifaceted dangers necessitates the implementation of innovative and adaptable strategies. The findings outlined in this article underscore the potential efficacy of harnessing AI and ML as pivotal instruments in counteracting the dissemination of extremist ideologies and activities within the digital realm. The escalating apprehension pertains to the expansion of extremist entities employing digital platforms for recruitment, propagandistic dissemination, and operational coordination. The recognition of the imperative to proactively confront this predicament is evident through collaborative initiatives, exemplified by the Global Internet Forum to Counter Terrorism, alongside the integration of AI-driven content moderation approaches by platforms such as Facebook and Twitter. The paradox of AI, wherein technology devised for benevolent objectives could be exploited by extremist factions to advance their agendas, encapsulates what is termed as the "dual-use dilemma" of artificial intelligence.

This study has endeavored to explore diverse applications of AI that can be employed to counteract online extremism. Methodologies encompassing automated content moderation, sentiment analysis, and predictive analytics assume increasing significance in monitoring, evaluating, and responding to content with potential extremist inclinations. Underpinned by AI, these technologies empower law enforcement and cyber security entities to anticipate threats, discern patterns of radicalization, and promptly intervene to avert plausible violent actions.

Against the backdrop of Pakistan's protracted struggle against extremism and terrorism, the nation's context assumes pronounced relevance within this discourse. The country contends with challenges posed by terrorist groups such as Daesh-K and TTP, which have demonstrated remarkable adaptability in acclimatizing to evolving internet platforms. As per the findings of this research, extremist organizations frequently exploit encrypted messaging apps and the dark web to sustain operations beyond the reach of conventional content monitoring mechanisms. The potential application of AI technologies customized to Pakistan's unique circumstances and linguistic diversity emerges as pivotal in monitoring and counteracting such activities. Nevertheless, at the same time it also is imperative to acknowledge the limitations inherent in the utilization of artificial intelligence within this context. AI is not an infallible solution, at least not yet, and confronts intrinsic issues including linguistic nuances, cultural variations, and the potential for false positives.

Bibliography

Ahmad, Shakeel, Muhammad Zubair Asghar, Fahad M. Alotaibi, and Irfanullah Awan. "Detection and Classification of Social Media-Based Extremist Affiliations Using Sentiment Analysis Techniques." Human-Centric Computing and Information Sciences 9, no. 1 (July 1, 2019): 24. https://doi.org/10.1186/s13673-019-0185-6.

Basheer, Randa, and Bassel Alkhatib. "Threats from the Dark: A Review over Dark Web Investigation Research for Cyber Threat Intelligence." Journal of Computer Networks and Communications 2021 (December 20, 2021): 1–21. https://doi.org/10.1155/2021/1302999.

Basit Khan, Abdul. "Creating Social Firewalls against Militants' Online Propaganda Is Critical for Pakistan's Internal Security." News and Analysis. Arab News PK, March 3, 2023. https://arab.news/vxv6p.

Bates, David W., Andrew Auerbach, Peter Schulam, Adam Wright, and Suchi Saria. "Reporting and Implementing Interventions Involving Machine Learning and Artificial Intelligence." Annals of Internal Medicine 172, no. 11 Suppl (June 2, 2020): S137–44. https://doi.org/10.7326/M19-0872.

Chen, Hsinchun. "Sentiment and Affect Analysis of Dark Web Forums: Measuring Radicalization on the Internet." In 2008 IEEE International Conference on Intelligence and Security Informatics, 104–9. Taipei, Taiwan: IEEE, 2008. https://doi.org/10.1109/ISI.2008.4565038.

Comiter, Marcus. "Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It." News&Analysis. Belfer Center for Science and International Affairs, August 2019. https://www.belfercenter.org/publication/AttackingAI.

Corbeil, Alexander, and Rafal Rohozinski. "Managing Risk: Terrorism, Violent Extremism, and Anti-Democratic Tendencies in the Digital Space." In The Oxford Handbook of Cyber Security, edited by Paul Cornish, 0. Oxford University Press, 2021. https://doi.org/10.1093/oxfordhb/9780198800682.013.8.

De Azambuja, Antonio João Gonçalves, Christian Plesker, Klaus Schützer, Reiner Anderl, Benjamin Schleich, and Vilson Rosa Almeida. "Artificial Intelligence-Based Cyber Security in the Context of Industry 4.0—A Survey." Electronics 12, no. 8 (April 19, 2023): 1920. https://doi.org/10.3390/electronics12081920.

Faiella, Mario, Gustavo Granadillo, Ibéria Medeiros, Rui Azevedo, and Susana Gonzalez-Zarzosa. "Enriching Threat Intelligence Platforms Capabilities," 2019. https://doi.org/10.5220/0007830400370048.

Fishman, Brian. "Dual-Use Regulation: Managing Hate and Terrorism Online before and after Section 230 Reform." Think Tank Website. Brookings, March 14, 2023. https://www.brookings.edu/articles/dual-use-regulation-managing-hate-and-terrorism-online-before-and-after-section-230-reform/

Giordano, James, ed. Neurotechnology in National Security and Defense: Practical Considerations, Neuroethical Concerns. 0 ed. CRC Press, 2014. https://doi.org/10.1201/b17454.

Gobble, MaryAnne M. "The Road to Artificial General Intelligence." Research-Technology Management 62, no. 3 (May 4, 2019): 55–59. https://doi.org/10.1080/08956308.2019.1587336.

Hao, Karen. "How Facebook Got Addicted to Spreading Misinformation." Technology Website. MIT Technology Review, March 11, 2021. https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/.

House, White. "Empowering Local Partners to Prevent Violent Extremism in the United States." Home Land Security Report, August 2011.

Bibliography

Ibrahim, Yasmin. “The Sharing Economy and Livestreaming of Terror: Co-Production of Terrorism on Social Media.” In Terrorism, Violent Radicalisation, and Mental Health, edited by Kamaldeep Bhui, Dinesh Bhugra, Kamaldeep Bhui, and Dinesh Bhugra, 0. Oxford University Press, 2021. https://doi.org/10.1093/med/9780198845706.003.0006.

Kane, Scott M. Martin, James R. Casey, Stephanie. “History of Artificial Intelligence and Personalized Learning.” In Serious Games in Personalized Learning. Routledge, 2021.

Kaur, Armaan, Jaspal Kaur Saini, and Divya Bansal. “Detecting Radical Text over Online Media Using Deep Learning.” arXiv, July 30, 2019. http://arxiv.org/abs/1907.12368.

Library of Congress, Washington, D.C. 20540 USA. “Germany: Network Enforcement Act Amended to Better Fight Online Hate Speech.” Web page. Accessed August 27, 2023. https://www.loc.gov/item/global-legal-monitor/2021-07-06/germany-network-enforcement-act-amended-to-better-fight-online-hate-speech/.

Lupovici, Amir. “The Dual-Use Security Dilemma and the Social Construction of Insecurity.” Contemporary Security Policy 42, no. 3 (July 3, 2021): 257–85. https://doi.org/10.1080/13523260.2020.1866845.

Munden, Hanna Paalgard. “Extremist Group Exits: What Autobiographies by Male Right-Wing Formers Reveal about Identity Transformation.” Behavioral Sciences of Terrorism and Political Aggression 0, no. 0 (April 3, 2023): 1–22. https://doi.org/10.1080/19434472.2023.2192771.

Ozmen Garibay, Ozlem, Brent Winslow, Salvatore Andolina, Margherita Antona, Anja Bodenschatz, Constantinos Coursaris, Gregory Falco, et al. “Six Human-Centered Artificial Intelligence Grand Challenges.” International Journal of Human–Computer Interaction 39, no. 3 (February 7, 2023): 391–437. https://doi.org/10.1080/10447318.2022.2153320.

Park, Chang Sup, and Homero Gil De Zúñiga. “Learning about Politics from Mass Media and Social Media: Moderating Roles of Press Freedom and Public Service Broadcasting in 11 Countries.” International Journal of Public Opinion Research 33, no. 2 (August 17, 2021): 315–35. https://doi.org/10.1093/ijpor/edaa021.

Persily, Nathaniel. “Platform Power, Online Speech, and the Search for New Constitutional Categories.” In Social Media, Freedom of Speech, and the Future of Our Democracy, edited by Lee C. Bollinger and Geoffrey R. Stone, 0. Oxford University Press, 2022. https://doi.org/10.1093/oso/9780197621080.003.0012.

Rainie, Lee. “Code-Dependent: Pros and Cons of the Algorithm Age.” Pew Research Center: Internet, Science & Tech (blog), February 8, 2017. https://www.pewresearch.org/internet/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/.

Rajawat, Anand Singh, Pradeep Bedi, S. B. Goyal, Sandeep Kautish, Zhang Xihua, Hanan Aljuaid, and Ali Wagdy Mohamed. “Dark Web Data Classification Using Neural Network.” Computational Intelligence and Neuroscience 2022 (March 28, 2022): 8393318. https://doi.org/10.1155/2022/8393318.

Recupero, Patricia R., and Samara E. Rainey. “The Internet and Social Media as an Enabling Force.” In Lone-Actor Terrorism: An Integrated Framework, edited by Andrew McCabe, John Wyman, Jacob C. Holzer, Andrea J. Dew, Patricia R. Recupero, and Paul Gill, 0. Oxford University Press, 2022. https://doi.org/10.1093/med/9780190929794.003.0011.

Rigano, Christopher. “Using Artificial Intelligence to Address Criminal Justice Needs.” Government Website. National Institute of Justice, October 8, 2018. https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs.

Bibliography

Scrivens, Ryan, Thomas W. Wojciechowski, Joshua D. Freilich, Steven M. Chermak, and Richard Frank. "Comparing the Online Posting Behaviors of Violent and Non-Violent Right-Wing Extremists." Terrorism and Political Violence 35, no. 1 (January 2, 2023): 192–209. https://doi.org/10.1080/09546553.2021.1891893.

Stackpole, Thomas. "Content Moderation Is Terrible by Design." Harvard Business Review, November 9, 2022. https://hbr.org/2022/11/content-moderation-is-terrible-by-design.

Weinbaum, Marvin G. "Insurgency and Violent Extremism in Pakistan." Small Wars & Insurgencies 28, no. 1 (January 2, 2017): 34–56. https://doi.org/10.1080/09592318.2016.1266130.

Williams, Thomas James Vaughan, Calli Tzani, Helen Gavin, and Maria Ioannou. "Policy vs Reality: Comparing the Policies of Social Media Sites and Users' Experiences, in the Context of Exposure to Extremist Content." Behavioral Sciences of Terrorism and Political Aggression 0, no. 0 (April 8, 2023): 1–18. https://doi.org/10.1080/19434472.2023.2195466.

Young, Greyson K. "How Much Is Too Much: The Difficulties of Social Media Content Moderation." Information & Communications Technology Law 31, no. 1 (January 2, 2022): 1–16. https://doi.org/10.1080/13600834.2021.1905593.

Zhou, Guangyou, Kuangxiong Gong, Sumei Luo, and Guohu Xu. "Inclusive Finance, Human Capital and Regional Economic Growth in China.