Journal of Infrastructure, Policy and Development 2024, 8(8), 6632.
https://doi.org/10.24294/jipd.v8i8.6632

Review

Social media algorithms in countering cyber extremism: A systematic review

Khalaf Tahat[1,2,*], Mohammed Habes[3], Ahmed Mansoori[1], Noura Naqbi[4], Najia Al Ketbi[5], Ihsan Maysari[6], Dina Tahat[7], Abdulaziz Altawil[1]

[1] Media & Creative Industries Department, United Arab Emirates University, Al Ain 15551, United Arab Emirates
[2] Journalism Department, Yarmouk University, Irbid 21163, Jordan
[3] Department of Radio & TV, Yarmouk University, Irbid 21163, Jordan
[4] Emirates College for Advanced Education, Abu Dhabi 126662, United Arab Emirates
[5] Mohammed Bin Zayed University for Humanities, Abu Dhabi 106621, United Arab Emirates
[6] Emirates News Agency, Abu Dhabi 00000, United Arab Emirates
[7] College of Education, Al Ain University, Al Ain 64141, United Arab Emirates
*** Corresponding author:** Khalaf Tahat, k.tahat@uaeu.ac.ae

CITATION

Tahat K, Habes M, Mansoori A, et al. (2024). Social media algorithms in countering cyber extremism: A systematic review. Journal of Infrastructure, Policy and Development. 8(8): 6632. https://doi.org/10.24294/jipd.v8i8.6632

ARTICLE INFO

Received: 24 May 2024
Accepted: 28 June 2024
Available online: 29 August 2024

Abstract: Countering cyber extremism is a crucial challenge in the digital age. Social media algorithms, if designed and used properly, have the potential to be a powerful tool in this fight, development of technological solutions that can make social networks a safer and healthier space for all users. this study mainly aims to provide a comprehensive view of the role played by the algorithms of social networking sites in countering electronic extremism, and clarifying the expected ease of use by programmers in limiting the dissemination of extremist data. Additionally, to analyzing the intended benefit in controlling and organizing digital content for users from all societal groups. Through the systematic review tool, a variety of previous literature related to the applications of algorithms in the field of online radicalization reduction was evaluated. Algorithms use machine learning and analysis of text and images to detect content that may be harmful, hateful, or call for violence. Posts, comments, photos and videos are analyzed to detect any signs of extremism. Algorithms also contribute to enhancing content that promotes positive values, tolerance and understanding between individuals, which reduces the impact of extremist content. Algorithms are also constantly updated to be able to discover new methods used by extremists to spread their ideas and avoid detection. The results indicate that it is possible to make the most of these algorithms and use them to enhance electronic security and reduce digital threats.

Keywords: social media; algorithms; cyber extremism; new media; artificial intelligence (AI)

1. Introduction

Technological progress and prosperity during the end of the last century made a quantum leap in the world of communications, especially with the emergence of social media and its invasion of all spectrums of societies (Monti et al., 2019), which later became an integral part of daily activities and a world without borders of science and knowledge as a result of openness to different cultures and societies (Jwaniat et al., 2023). Recently, the role of these social interactive platforms has grown significantly until they are no longer just secondary tools in entertainment, but rather have gone beyond being a method used for daily communication between individuals to become a reliable reference by its users to receive information and in more than one way, depending on Due to the different policies for presenting it from one side to the other (Shah, 2018). It should be noted here the importance and effectiveness of social networking sites because of their ease and speed in obtaining information (Monti et

al., 2019) and referring to them at any time and their ability to influence individuals by supporting publications using various multimedia (Salem, 2021). In addition to the audience's ability to interact with it and express it however they want and whenever they want (Montasari et al., 2022).

The Internet facilitates connection and cooperation between different extremist groups worldwide. This can lead to the formation of dangerous alliances and the transfer of knowledge and resources between them (Bondad-Brown et al., 2012). Tackling cyber extremism requires close collaboration between governments, technology companies and civil society. (Agarwal and Sureka, 2015). It is essential to develop effective prevention and response strategies, as well as educate the public about the risks and signs of online radicalization, only through a coordinated and multifaceted effort can this growing threat be mitigated and global security and stability protected (Monti et al., 2019).

In today's digital age, social media has radically transformed the way people communicate and access information. However, these platforms have also become fertile ground for the spread of cyber extremism (Bondad-Brown et al., 2012). This phenomenon involves the use of the Internet to promote extremist ideologies, recruit followers, and coordinate harmful activities (Awan, 2017). Cyber extremism not only threatens global security, but also erodes social cohesion and fuels hatred and violence (Awan, 2017). In light of the growing reliance of users on social networking sites to obtain information (Martínez-López et al., 2022) it was used as a promotional platform to spread extremist opinions and trends that express the ideologies of individuals or groups by employing various multimedia (Agarwal and Sureka, 2015) which necessitated the need to find a framework for the publications from moral violations that may lead the individual to violations and legal consequences, at a time when users find it platforms for expressing opinions and communicating with extremists and terrorists without any limits and controls (Ahmad et al., 2019). Since social networks have proven their ability to form intellectual public opinion trends (Al-Garadi et al., 2019) artificial intelligence technologies have launched with their tremendous power, and the imposition of a new reality on social networking sites in benefiting from them by directing them towards accurate social analyzes that infer Innovative solutions, better planning, and faster sharing of knowledge (Alzubi et al., 2018). It was found that the automated management of the content of social media platforms can limit the spread of extremist content (Weimann and Am, 2020) through what the algorithms of social networking sites were trained on (Amit et al., 2021) as it had an important role and artificial intelligence in mediating users' electronic interactions (Badr et al., 2019). This raises an area for research to find out the role of social networking algorithms in countering extremism (Benabdelouahed and Dakouan, 2020).

2. Statement of problem

Studies indicate that various social media platforms have become a haven for extremism and extremist groups seeking to spread their extremist ideologies and beliefs in order to spread their agendas, which they share with virtual communities created via the Internet (Berberich et al., 2020; Bergström and Jervelycke Belfrage, 2018; Bouko et al., 2021). Social media has become one of the important tools in

changing behavior, attitudes, and ideas affecting intellectual extremism in individuals as a result of their addiction to social networks (Dencik et al., 2015). Many scholars assert that the Internet has succeeded as a "facilitating environment" that causes violent extremism (Duarte et al., 2017). In terms of the prevalence of social media for all users, artificial intelligence techniques appeared (Kurniawan and Surendro, 2018), which worked to identify and ban extremism on social media (Leonardi, 2021) by analyzing and interpreting data and human inputs, which works to process them and produce outputs, in while the operations carried out by the algorithms are not limited to confronting extremism, but were developed to go beyond it to predict the time when extremism and hate crimes are likely to occur (Mahesh, 2020) and this is what makes Different organizations resort to using algorithms. From this standpoint, the problem of the study lies in answering the following main question: What is the role of social networking algorithms in confronting electronic extremism?

3. Theoretical literature

The step of reviewing the theoretical literature related to the research problem is an important step when conducting studies (Habes et al., 2023b; Mathur et al., 2018).

3.1. Social networking sites are the fields of spreading extremist ideologies

Through previous studies, it was observed that many individuals use social networking sites such as Facebook for personal motives such as communication (Mazza et al., 2017), education (Singh and Goraya, 2019), obtaining information (Micu et al., 2018), and other motives. But it went beyond that to become also a fertile ground for extremist political groups on a large scale, especially on the Facebook platform in relative secrecy, who found themselves on the margins of their societies. They enacted their extremist beliefs and ideas, seeking through these platforms to spread their discourse and recruit new followers, as it facilitated the path of contacting and reaching with like-minded people (Monti et al., 2019).

For example, social media, online forums were used to spread his extremist ideology globally. Through propaganda videos and social media posts, they managed to recruit thousands of foreign fighters from various countries (Ehteshami Bejnordi et al., 2017). Platforms such as 8chan have been used by right-wing extremists to radicalize individuals. The Christchurch attacker in New Zealand, who killed 51 people at two mosques in 2019, had shared his manifesto on these forums before the attack. In 2015, hackers linked to Russian extremist groups carried out a cyberattack that left nearby power out. of 230,000 people in Ukraine (Petrescu and Krishen, 2020). This attack demonstrated the potential of cyberattacks to wreak havoc on critical infrastructure (Davis, 2021). In 2016, Russian entities were found to have used social media to spread disinformation and extremist propaganda during the US presidential election. This included the creation of fake accounts that promoted social and political divisions (Tuteja and Marwaha, 2023)

Telegram algorithms

Telegram is a cloud-based instant messaging app known for its strong emphasis on privacy and security. Telegram offers end-to-end encryption for its "Secret Chats,"

ensuring that messages can only be read by the intended recipients (Davis, 2021; Ionescu et al., 2020; Tuteja and Marwaha, 2023). This makes it difficult for authorities to monitor communications (Ionescu et al., 2020). Unlike traditional social media platforms where content is publicly visible, Telegram uses channels and groups. Channels can broadcast messages to an unlimited number of subscribers, while groups can host up to 200,000 members (Walther and McCoy, 2021). These features allow for the rapid dissemination of information to large audiences without public scrutiny (Tuteja and Marwaha, 2023). As well as Telegram has been criticized for its lenient content moderation policies. Unlike platforms like Facebook and Twitter, which have more stringent guidelines and actively remove extremist content, Telegram is known for allowing a wider range of speech. This makes it a preferred platform for extremist groups who face bans on other social media sites (Davis, 2021).

In light of the proliferation of those websites that allow individuals to express their opinions and try to obtain their violated rights, through which they are unable to differentiate between good and bad ideas, especially since these social media include forums, chat rooms, and groups, which makes many young people adopt extremist ideas, especially since it does not affect them selectively (Biswal and Gouda, 2020). The extremist propaganda movement, who waged informational and psychological wars on social networking sites, became active in service of their traditional wars (Mullah and Zainon, 2021), as it was found that the Facebook platform is the most used platform by extremists to spread their extremist ideas and that Telegram allow extremists to arrange themselves into attractive virtual groups that reflect the structure of their physical counterparts, while the Internet has been considered a hallmark of modern patterns of extremism (Pauwels et al., 2014). Extremists found it useful in spreading their extremist beliefs and ideas due to the ease of creating and interacting with content, as well as its ability to form a large user base that supports anonymity (Petrescu and Krishen, 2020). Furthermore, Telegram does not participate in the comprehensive content removal policies compared to other platforms and only removes pornography and some violent speech from its public channels (Saveliev and Zhurenkov, 2021) While platforms such as YouTube facilitate forms of popularity, fame, and harmony among extremist content producers, which lends legitimacy and credibility to them (Shah and Jha, 2018)

The continuous developments of the communications and technology revolution enabled the receiver to become a transmitter at the same time (Sharma et al., 2020), there are millions of fake pages and profiles that seek to spread misinformation that will spread hatred, but in the midst of all this, many local, regional and international institutions are also working to counter extremism by educating users and presenting counter-narratives that intercepts assumptions and shows fallacies by employing various multimedia (Srivastava et al., 2021), thus confronting extremism with the same weapon (Trivedi and Singh, 2017).

From 2005 to 2016, the use of social media platforms by extremists to spread their ideas evolved significantly. This period saw the rise of various social media networks, which provided extremists with new avenues to disseminate their ideologies, recruit members, and coordinate activities. The decentralized and often anonymized nature of these platforms allowed extremist content to flourish, challenging traditional methods of monitoring and control. During this period, the use of social media by

extremists became more sophisticated and widespread. Platforms like Twitter and Facebook became central to their strategies by 60% (Sharma and Sicinski, 2020). The anonymity and global reach provided by these platforms allowed them to spread their messages more effectively than ever before. As a result, monitoring and countering extremist content on social media has become a critical challenge for governments and tech companies alike (Chacko et al., 2016).

3.2. Social networking algorithms

Algorithms are the product of the modern technological revolution in general, and artificial intelligence in particular, and they are considered the cornerstone on which social media was built until they are used in various technological activities such as marketing and electronic selling (Valentini et al., 2020), detecting threats on social media (Waldman and Verga, 2016), and detecting fake news (Walther and McCoy, 2021). Explains the concept of algorithms as "coded procedures for converting input data into desired output" or "codes that automate tasks" as they act as "semi-rhetorical agents" due to the nature of their performance and the beliefs included and encoded in their structure and overall structure. It has recently become increasingly noticeable in the processing of big data by understanding past events and predicting future behavior facilitating the possibilities of taking preventive action (Walther and McCoy, 2021; Yadav et al., 2020). These algorithms work on several issues that would improve human behavior on social media platforms in general and Facebook in particular (Trivedi and Singh, 2017) but in one way or another they lack emotional intelligence (Yadav et al., 2020).

In May 2021, Facebook created its own Transparency Center and regularly updates its Community Standards, which are reviewed by the Facebook Oversight Board according to concrete cases published since January 2021 (Sharma et al., 2020). In this regard, Facebook has witnessed many developments in its algorithms since its inception, the most prominent of which was the ability to choose what the user wanted to see or hide from ads; so that it then gives the user a set of options to determine the reason for his desire for the advertisement not to appear to him, and this also applies to publications and public pages on it, which helps to determine the type of content that the user wishes to be exposed to or that he prefers to avoid, and from it, Facebook deletes the content that opposes its community standards as well It can also remove or restrict audiences for certain types of sensitive content such as nudity, violence, etc. (Salem, 2021).

From 2004 to 2020, Facebook's algorithms evolved from a simple chronological feed to a sophisticated system designed to maximize user engagement, prioritize meaningful interactions, and combat misinformation (Martínez-López et al., 2022). These changes reflect Facebook's efforts to balance user satisfaction with the business interests of advertisers, while also addressing the growing concerns about the impact of social media on society. The platform's continued focus on transparency and user control aims to maintain trust and relevance in an ever-changing digital landscape.

In the era of big data on social networking sites, algorithms try to analyze and examine it with a high accuracy that human brains cannot visualize or do (Agarwal and Sureka, 2015). However, the increasing consumption of big data comes within the

framework of serious data protection and civil rights issues such as the right to privacy, especially since it lacks accountability and transparency (Thomas-Evans, 2022) as the data obtained can be sold to another party for various purposes (Al-Garadi et al., 2019), but it can be argued that being the target of hidden algorithmic analysis is an unavoidable part of using major Internet services and social networks and is the price we pay for services provided for free (Bergström and Jervelycke Belfrage, 2018). The main problem related to algorithmic techniques on social media remains centered on how to ensure its objectivity and accuracy in information handling, despite being subject to standards, inputs, and outputs that may ensure impartiality when programming them (Amit et al., 2021).

3.3. Machine Learning (ML) algorithms on social media in the face of extremism

Mahesh (2020) describes machine learning (ML) as "the scientific study of algorithms and statistical models that computer systems use to perform a specific task without being explicitly programmed" (Valentini et al., 2020). Machine learning algorithms help researchers understand big data, as it is possible through its application on social media to obtain the largest amount of information available to be exploited in detecting negative practices on it (Sharma et al., 2020).

Alzubi et al. (2018) showed that data processing on social media can only take place if the problem is properly classified, so that the most appropriate machine learning algorithm is applied to it and any problem in data science can be grouped into one of the five categories.

Mullah and Zainon (2021) illustrated the different types of problems that can be addressed using machine learning techniques. Each type of problem requires specific approaches and algorithms to effectively analyze data and generate insights. The exploration of different types of machine learning problems highlights the diverse applications and methodologies that define this field. As illustrated by Mullah and Zainon (2021), machine learning problems can be broadly categorized into Classification, Anomaly Detection, Regression, Clustering, and Reinforcement Learning problems. Each of these categories addresses unique challenges and requires specific algorithms and techniques to generate accurate and meaningful insights (Alzubi et al., 2018).

In this context, the researchers noticed that the issue of classifications is considered one of the most important issues in the science of machine learning, so they took care of it and invented many algorithms that contribute to addressing many problems. Among these algorithms is the (SVM) algorithm, which is the abbreviation for the term (Support Vector Machine), as it was considered the best choice for classifying and analyzing texts on social networking sites (Mullah and Zainon, 2021; Petrescu and Krishen, 2020; Saveliev and Zhurenkov, 2021) specifically, the classification of hate speech or not, followed in order of efficiency by the (RF) algorithm (Random Forest), and the Logistic Regress (LR) algorithm, then (Naive Bayes (NB)) is also used well (Leonardi, 2021).

In the context of hate speech detection, selecting the appropriate algorithm is crucial for achieving high accuracy and efficiency. Random Forest (RF) is the most

efficient due to its robustness and ability to handle complex data structures. Logistic Regression (LR) offers a balance between efficiency and interpretability, making it suitable for scenarios where understanding the model's decisions is important. Naive Bayes (NB), while less accurate in some cases, provides a fast and effective solution, especially as a baseline model (Bdoor and Habes, 2024; Ullah et al., 2024). The comparative analysis presented by Leonardi (2021) underscores the importance of algorithm selection in hate speech classification tasks. By leveraging the strengths of these algorithms, practitioners can effectively detect and mitigate the spread of hate speech on digital platforms, contributing to a safer and more inclusive online environment (Bdoor and Habes, 2024; Ullah et al., 2024).

Experiments were conducted with deep learning algorithms on various applications to train the classification of publications, and through them it was found that the performance is better for the convolutional neural network (CNN) in a data set in addition to the classification of texts that incite hatred in textual content on the Internet in languages such as Arabic and Vietnamese (Srivastava et al., 2021), as well as recurrent neural network (RNN) because they capture sentence semantics better (Amit et al., 2021).

The use of machine learning algorithms (Machine Learning, ML) on social networks to combat extremism is a practice that has gained relevance in recent years (Salloum et al., 2023). These algorithms make it possible to detect, analyze and mitigate the spread of extremist content more efficiently than manual methods, as well as using machine learning algorithms on social media to combat extremism is a powerful and necessary tool, but it must be handled carefully (Bouko et al., 2021). Thus the combination of advanced technologies with human supervision and constant attention to ethical and legal aspects is crucial to effectively and responsibly address this global problem (Amit et al., 2021).

3.4. Perceived benefit of algorithms in countering extremism

The perceived benefit lies in the extent to which individuals should expect that the way they use the new technology enhances their performance (Alzubi et al., 2018). It can be said here that artificial intelligence techniques, which are defined as "the science of making machines smart", work to develop systems in line with the intellectual characteristics of humans. Within the framework of community service, Instagram and Facebook adopt artificial intelligence techniques to delete fake messages from accounts that may indulge in extremism by training them to solve problems, whether textual or contained in images, through which it allows identifying all elements in the image and then dealing with them based on algorithms equipped to identify Images through the camera system (Ahmad et al., 2019). Artificial intelligence techniques and algorithms can strategically monitor digital platforms (Bouko et al., 2021) and perform predictive analyzes for them that emerging information technology infrastructures based on machine learning algorithms contribute to their development (Mullah and Zainon, 2021), in addition to accessing the huge amount of social media data, and the development of algorithms for vetting posts gives police agencies the ability to expand and digitize proactive policing strategies (Waldman and Verga, 2016). Moreover, the automated management of

content helps limit and stop the spread of terrorist content, as many social media companies are keen to prevent malicious actors from exploiting their platforms, as machine learning algorithms play their role by filtering 98% of harmful content on Facebook, and according to what Twitter announced that it deals with ten accounts per second. In turn, Google removes 80% of inappropriate videos before anyone watches them. According to Mark Zuckerberg, CEO of Facebook, automated systems can only process the content of one million users in different languages (Shah and Jha, 2018).

In the context of countering digital extremism, faculty and students at Columbia University's College of Social Work and Data Science Institute launched the Science Computer and Intervention Gang project, a research study that uses publicly available Twitter data from youth involved in gangs in Chicago to illustrate the conditions that constitute aggressive and threatening online communications, and developing computational tools to stop incited violence through communication through social media (Shah and Jha, 2018). In the midst of countering digital extremism (Trivedi and Singh, 2017) called for more research on frameworks that would enhance algorithmic models for social media platforms (Saveliev and Zhurenkov, 2021).

The perceived ease of use of any given system is referred to as "the level of technology used with a perception of its proper use" (Walther and McCoy, 2021). Regardless of the field of digital media, AI technologies have proven successful in the fields in which they are studied. Then, the perceived ease of use has a great relationship with the behavioral intention to use, and based on the employment of artificial intelligence in digital media, it can be said that the perceived ease of use is the level at which programmers see that the use of artificial intelligence projects will not require much effort in combating extremism digitally because the task will be easier by using Algorithms.

As the study (Walther and McCoy, 2021) showed that these technologies can identify spelling errors in search engines and suggest alternatives, which may contribute to avoiding access to extremist data, which is now being targeted by extremists (Valentini et al., 2020). Countries such as the United States, Spain, Russia, and the United Kingdom have concluded that the violent attacks they encountered were facilitated by extremists on the Internet and social media have become one of the most growing concerns in recent times (Saveliev and Zhurenkov, 2021). However, Natural Language Processing (NLP) algorithms have been developed that can analyze SM language and make judgments about the types of content and related vocabularies and advanced SM monitoring techniques such as Neuro-Linguistic Programming (NLP) are being developed to monitor open source intelligence (OSINT) to facilitate prediction or detection of terrorist events (Waldman and Verga, 2016).

As Duarte and his team on 2018 shows, government agencies can purchase off-the-shelf natural language processing (NLP) tools designed for a range of purposes such as translating text, filtering offensive content, and improving spam detection with what they are trained on using examples of texts that humans have categorized as either belonging to a target category of content or not (Valentini et al., 2020). In addition to the content censorship imposed by social media platforms, which concluded that Twitter's deletion of accounts supporting the Islamic State, for example, "significantly affected its ability to create and maintain strong communities", and it ended up preventing many extremists from using social media by the Internet (Shah and Jha,

2018), who found the Internet their best option for its ease of access to each other, and to address a broad and global audience using a broad and dynamic range of narratives, which raised concerns of an increase in recruitment and violent extremism under the influence of the Internet (Pauwels et al., 2014).

The Global Internet Forum for Counter-Terrorism (GIFCT) worked on a database of "segmentation" to enable the rapid elimination of extremist connotation across platforms and websites, but all efforts to prevent extremism on social media were countered through what were called "tech alt" platforms, which was a platform for extremists banned from the main platforms as a safe place for extremist ideas (Mullah and Zainon, 2021).

Artificial intelligence techniques have enabled advanced organizations to examine their customers' practices and search for their tendencies on the Internet, which ultimately encourages reaching their intended marketing interests (Monti et al., 2019). In the context of analyzing the data obtained by these algorithms on social networks, Micu et al. (2018) show that the data spread on the Internet more easily and is used as a reference that may be misused, considering the work of others as their own property, intentionally or unintentionally. While both (Mazza et al., 2017) point out that there are developments in specialized standards, such as protecting personal data for users, Schroeter argues in his book Artificial Intelligence and Countering Violent Extremism that algorithms need huge amounts of data to make useful predictions about the future (Walther and McCoy, 2021). While that data also addressed the use of social media by extremists, social media companies have ramped up the use of artificial intelligence technologies to address the promotion of extremist content online by identifying and removing it from websites at a faster pace and with higher accuracy, through artificial intelligence techniques Facebook removed 837 million spam and 5.2 million content promoting hate speech and deactivated 583 million accounts globally during the first quarter of 2018 alone. Also, within 23 months, starting in August 2015, Twitter suspended about a million accounts promoting violence and in the latter half of 2017, YouTube removed 150,000 videos promoting extremism, with about half of those videos removed within two hours of being uploaded (Valentini et al., 2020).

On the other hand, social networking sites are viewed as mere platforms and private technology companies that are not obligated to extend the protection of the content that their users share on their platforms, as they are not media companies (Mansoori et al., 2023; Tahat et al., 2023b). This is what gives social media tremendous power to influence public discourse. For example, the policies of Facebook and Instagram, inspired only by the First Amendment to the US Constitution related to freedom of expression including hate speech as long as it is not based on fighting words, incitement to violence, or real threats (Mullah and Zainon, 2021).

4. Conclusion

Algorithms can play a significant role in both the spread and the prevention of extremism through digital media. On the one hand, algorithms can be used to amplify extremist content, making it more visible to users and increasing the risk of radicalization. On the other hand, algorithms can also be used to combat extremism. This study presented an attempt to clarify the mechanism of digital extremism

reduction through a systematic review of previous literature on this subject. Based on this, it can be said that in light of the widespread use of social media algorithms in organizing many digital operations, their ability to limit the spread of electronic extremism comes from the possibility of analyzing huge data and content on the Internet in order to find suspicious patterns and irresponsible digital behaviors. By taking advantage of machine learning techniques and natural language analysis, phrases related to extremism can be identified and their content deleted. In addition, the ability of algorithms to predict human behavior constitutes an important barrier in countering extremism before it spreads widely. This helps maintain a secure digital environment. The role of algorithms in combating extremism through digital media is still evolving. There is no single solution that will work in all cases, and it is important to use algorithms carefully and responsibly. However, if used effectively, algorithms can be a powerful tool in the fight against extremism. Based on what the study clarified and concluded, it recommends researchers to research the forms of camouflage and fraud that extremists adopt in spreading their ideologies and the mechanism of hunting them by algorithms after developing and training them in this.

Author contributions: Conceptualization, IM and NAK; methodology, KT; software, NAK; validation, AA, IM and NAK; formal analysis, KT; investigation, KT and MH; resources, MH and NAK; writing—original draft preparation, AA, MH and IM; writing—review and editing, AM and NN; supervision, DT. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

Agarwal, S., & Sureka, A. (2015). Applying social media intelligence for predicting and identifying on-line radicalization and civil unrest oriented threats. arXiv. arXiv:1511.06858.

Ahmad, S., Asghar, M. Z., Alotaibi, F. M., et al. (2019). Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. Human-Centric Computing and Information Sciences, 9(1). https://doi.org/10.1186/s13673-019-0185-6

Al-Garadi, M. A., Hussain, M. R., Khan, N., et al. (2019). Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges. IEEE Access, 7, 70701–70718. https://doi.org/10.1109/access.2019.2918354

Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine learning from theory to algorithms: an overview. Journal of Physics: Conference Series, 1142, 012012. https://doi.org/10.1088/1742-6596/1142/1/012012

Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine Learning from Theory to Algorithms: An Overview. Journal of Physics: Conference Series, 1142, 012012. https://doi.org/10.1088/1742-6596/1142/1/012012

Amit, S., Barua, L., & Kafy, A.-A. (2021). Countering violent extremism using social media and preventing implementable strategies for Bangladesh. Heliyon, 7(5), e07121. https://doi.org/10.1016/j.heliyon.2021.e07121

Awan, I. (2017). Cyber-Extremism: Isis and the Power of Social Media. Society, 54(2), 138–149. https://doi.org/10.1007/s12115-017-0114-0

Badr, E. M., Salam, M. A., Ali, M., & Ahmed, H. (2019). Social media sentiment analysis using machine learning and optimization techniques. International Journal of Computer Applications, 975, 8887.

Bdoor, S. Y., & Habes, M. (2024). Use Chat GPT in Media Content Production Digital Newsrooms Perspective. In: Artificial Intelligence in Education: The Power and Dangers of ChatGPT in the Classroom. Springer. (pp. 545–561)

Benabdelouahed, R., & Dakouan, C. (2020). The use of artificial intelligence in social media: opportunities and perspectives. Expert Journal of Marketing, 8(1), 82–87.

Berberich, N., Nishida, T., & Suzuki, S. (2020). Harmonizing artificial intelligence for social good. Philosophy & Technology, 33, 613–638.

Bergström, A., & Jervelycke Belfrage, M. (2018). News in social media: Incidental consumption and the role of opinion leaders. Digital Journalism, 6(5), 583–598.

Biswal, S. K., & Gouda, N. K. (2020). Artificial intelligence in journalism: A boon or bane? Optimization in Machine Learning and Applications, 155–167.

Bondad-Brown, B. A., Rice, R. E., & Pearce, K. E. (2012). Influences on TV viewing and online user-shared video use: Demographics, generations, contextual age, media use, motivations, and audience activity. Journal of Broadcasting & Electronic Media, 56(4), 471–493.

Bouko, C., Van Ostaeyen, P., & Voué, P. (2021). Facebook's policies against extremism: Ten years of struggle for more transparency. First Monday.

Chacko, A., Jensen, S. A., Lowry, L. S., et al. (2016). Engagement in behavioral parent training: Review of the literature and implications for practice. Clinical Child and Family Psychology Review, 19(3), 204–215.

Davis, A. L. (2021). Artificial Intelligence and the Fight Against International Terrorism. American Intelligence Journal, 38(2), 63–73.

Dencik, L., Hintz, A., Carey, Z., & Pandya, H. (2015). Managing 'threats': uses of social media for policing domestic extremism and disorder in the UK. Available online: https://orca.cardiff.ac.uk/id/eprint/85618/ (accessed on 9 March 2024).

Duarte, N., Llanso, E., & Loup, A. (2017). Mixed messages? The limits of automated social media content analysis. In: Proceedings of the 1st Conference on Fairness, Accountability and Transparency. PMLR, 81, 106-106.

Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., et al. (2017). Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA, 318(22), 2199. https://doi.org/10.1001/jama.2017.14585

Habes, M., Elareshi, M., Safori, A., et al. (2023b). Understanding Arab social TV viewers' perceptions of virtual reality acceptance. Cogent Social Sciences, 9(1). https://doi.org/10.1080/23311886.2023.2180145

Habes, M., Tahat, K., Tahat, D., et al. (2023c). The Theory of Planned Behavior Regarding Artificial Intelligence in Recommendations and Selection of YouTube News Content. 2023 International Conference on Multimedia Computing, Networking and Applications (MCNA). https://doi.org/10.1109/mcna59361.2023.10185878

Ionescu, B., Ghenescu, M., Rastoceanu, F., et al. (2020). Artificial Intelligence Fights Crime and Terrorism at a New Level. IEEE MultiMedia, 27(2), 55–61. https://doi.org/10.1109/mmul.2020.2994403

Jwaniat, M. A. (2023). Examining Journalistic Practices in Online Newspapers in the Era of Artificial Intelligence. 2023 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS). https://doi.org/10.1109/iccns58795.2023.10193607

Kurniawan, M. A., & Surendro, K. (2018). Similarity measurement algorithms of writing and image for plagiarism on Facebook's social media. IOP Conference Series: Materials Science and Engineering, 403, 012074. https://doi.org/10.1088/1757-899x/403/1/012074

Leonardi, P. M. (2020). COVID-19 and the New Technologies of Organizing: Digital Exhaust, Digital Footprints, and Artificial Intelligence in the Wake of Remote Work. Journal of Management Studies, 58(1), 249–253. Portico. https://doi.org/10.1111/joms.12648

Mahesh, B. (2020). Machine Learning Algorithms—A Review. International Journal of Science and Research (IJSR), 9(1), 381–386. https://doi.org/10.21275/art20203995

Mansoori, A., Tahat, K., Tahat, D., et al. (2023). Gender as a moderating variable in online misinformation acceptance during COVID-19. Heliyon, 9(9), e19425. https://doi.org/10.1016/j.heliyon.2023.e19425

Martínez-López, F. J., Li, Y., & Young, S. M. (2022). Social Media Monetization. In Future of Business and Finance. Springer International Publishing. https://doi.org/10.1007/978-3-031-14575-9

Mathur, R., Bandil, D., & Pathak, V. (2018). Analyzing sentiment of twitter data using machine learning algorithm. Journal of Inventions in Computer Science and Communication Technology, 4(2), 1–7.

Mazza, C., Monaci, S., & Taddeo, G. (2017). Designing a social media strategy against violent extremism propaganda: The# heartofdarkness campaign. Available online: https://iris.polito.it/handle/11583/2700939 (accessed on 9 March 2024).

Micu, A., Capatina, A., & Micu, A.-E. (2018). Exploring artificial intelligence techniques' applicability in social media marketing. Journal of Emerging Trends in Marketing and Management, 1(1), 156–165.

Montasari, R., Carroll, F., Mitchell, I., et al. (2022). Privacy, Security And Forensics in The Internet of Things (IoT). Springer International Publishing. https://doi.org/10.1007/978-3-030-91218-5

Monti, F., Frasca, F., Eynard, D., et al. (2019). Fake news detection on social media using geometric deep learning. arXiv. arXiv:1902.06673.

Mullah, N. S., & Zainon, W. M. N. W. (2021). Advances in machine learning algorithms for hate speech detection in social media: a review. IEEE Access, 9, 88364–88376.

Pauwels, L., Brion, F., & De Ruyver, B. (2014). Explaining and understanding the role of exposure to new social media on violent extremism. Academia Press.

Petrescu, M., & Krishen, A. S. (2020). The dilemma of social media algorithms and analytics. Journal of Marketing Analytics, 8, 187–188.

Salem, D. F. (2021). The effectiveness of using artificial intelligence techniques in social networking sites from the point of view of educational media students: Facebook as a model. Egyptian Journal of Public Opinion Research, 20(3), 1–61.

Salloum, S. A., Bettayeb, A., Salloum, A., et al. (2023). Novel machine learning based approach for analysing the adoption of metaverse in medical training: A UAE case study. Informatics in Medicine Unlocked, 101354.

Saveliev, A., & Zhurenkov, D. (2021). Artificial intelligence and social responsibility: the case of the artificial intelligence strategies in the United States, Russia, and China. Kybernetes, 50(3), 656–675.

Shah, N. R., & Jha, S. K. (2018). Exploring organisational understanding of foundational pillars of social media a qualitative content analysis of social media policies of technology companies. Journal of Management Research, 18(4), 226–245.

Shah, W. A. (2018). Media disregarding laws on privacy of sexual abuse victims. Available online: https://www.dawn.com/news/1387393 (accessed on 9 March 2024).

Sharma, K., Seo, S., Meng, C., et al. (2020). Covid-19 on social media: Analyzing misinformation in twitter conversations. arXiv. arXiv:2003.12309.

Sharma, S., & Sicinski, P. (2020). A kinase of many talents: non-neuronal functions of CDK5 in development and disease. Open Biology, 10(1), 190287.

Singh, J., & Goraya, M. S. (2019). Multi-objective hybrid optimization based dynamic resource management scheme for cloud computing environments. In: Proceedings of 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT). pp. 386–391.

Srivastava, A., Hasan, M., Yagnik, B., et al (2021). Role of artificial intelligence in detection of hateful speech for Hinglish data on social media. In: Applications of Artificial Intelligence and Machine Learning: Select Proceedings of ICAAAIML 2020. Springer. pp. 83–95.

Tahat, K., Tahat, D. N., Masoori, A., et al. (2023b). Role of Social Media in Changing the Social Life Patterns of Youth at UAE. In Artificial Intelligence (AI) and Finance. Springer. pp. 152–163.

Trivedi, N. K., & Singh, S. K. (2017). A Systematic Survey on Detection of Extremism in Social Media. International Journal of Research and Scientific Innovation (IJRSI), 4(7), 94–103.

Tuteja, V., & Marwaha, S. S. (2023). Artificial intelligence: threat of terrorism and need for better counter-terrorism efforts. International Journal of Creative Computing, 2(1), 87–100.

Ullah, N., Al-Rahmi, W. M., Alblehai, F., et al. (2024). Blockchain-Powered Grids: Paving the Way for a Sustainable and Efficient Future. Heliyon.

Valentini, D., Lorusso, A. M., & Stephan, A. (2020). Onlife extremism: Dynamic integration of digital and physical spaces in radicalization. Frontiers in Psychology, 11, 524.

Waldman, S., & Verga, S. (2016). Countering violent extremism on social media. Defence Research and Development Canada, 1, 1–28.

Walther, S., & McCoy, A. (2021). US extremism on Telegram. Perspectives on Terrorism, 15(2), 100–124.

Weimann, G., & Am, A. Ben. (2020). Digital dog whistles: The new online language of extremism. International Journal of Security Studies, 2(1), 4.

Yadav, B. P., Sheshikala, M., Swathi, N., et al. (2020). Women Wellbeing Assessment in Indian Metropolises Using Machine Learning models. IOP Conference Series: Materials Science and Engineering, 981(2), 22042.