ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202310.00726

Postprint: An Analysis of AI-Based Governance of Cyber Violence

Authors: Sun Shi

Date: 2023-10-08T00:00:00+00:00

Abstract

【Purpose】 With the rapid development of information technology, the ways in which human society communicates, entertains, and acquires information have been diversified, while also giving rise to various problems and challenges. The rational, effective, and standardized application of artificial intelligence technology is an important means for governing modern social issues and is crucial to the effective implementation of precision social governance. This article aims to explore the introduction of artificial intelligence technology to govern online violence; 【Method】 The article clarifies the concept, causes, and harms of online violence, combined with relevant overviews of artificial intelligence, and analyzes the feasibility and specific application cases of employing artificial intelligence technology for precision governance of online violence; 【Result】 Artificial intelligence technology can effectively intercept and block online violent language, images, and videos, and provide effective intervention, thereby reducing the occurrence of online violent behavior; 【Conclusion】 Artificial intelligence technology can significantly prevent the occurrence of online violence, and the application of artificial intelligence technology in governing online violence can promote innovation in public governance approaches.

Full Text

An Analysis of Governance of Cyber Violence Based on Artificial Intelligence Technology

Sun Shi
School of Public Administration, Guangdong University of Finance and Economics, Foshan, Guangdong 528100

Abstract: With the rapid development of information technology, human society' s modes of communication, entertainment, and information acquisition have become increasingly diversified, while simultaneously giving rise to various

problems and challenges. The rational, effective, and standardized application of artificial intelligence technology constitutes an important means for addressing modern social issues and is crucial to the effective implementation of precision social governance. This article aims to explore the introduction of AI technology for governing cyber violence. The article clarifies the concept, causes, and harms of cyber violence, and in conjunction with an overview of artificial intelligence, analyzes the feasibility and specific application cases of introducing AI technology for precision governance of cyber violence. The results demonstrate that AI technology can effectively intercept and block cyber violent language, images, and videos while providing effective intervention, thereby reducing the occurrence of cyber violence. The conclusion indicates that AI technology can substantially prevent cyber violence, and its application in governing cyber violence can promote innovation in public governance approaches.

Keywords: artificial intelligence; cyber violence; precision governance; public governance innovation; digital governance

Classification: G223
Document Code: A
Article ID: 1671-0134(2023)01-064-06
DOI: 10.19483/j.cnki.11-4653/n.2023.01.011

Citation Format: Sun Shi. An Analysis of Governance of Cyber Violence Based on Artificial Intelligence Technology[J]. China Media Technology, 2023(01): 64-68, 87.

1. Problem Statement

The rapid development of contemporary information technology represented by the Internet has made human social life more convenient and diversified. However, it has also given rise to entirely new social problems, with various social issues occurring frequently and showing trends of intensification and diffusion, thereby increasing pressure on public governance. This situation urgently requires innovation in governance theories and improvements in governance capabilities, while also fostering innovative thinking about using information technology to solve public problems. Cyber violence, emerging from the rapid development of social networks, represents the most typical social phenomenon of our time. Its real-time, concentrated, and direct harm has caused many people to suffer deeply. Consequently, exploring more solutions to cyber violence has become imperative.

2. Overview of Artificial Intelligence

Intelligence arising through natural evolution and development is called natural intelligence. Human intelligence is currently the most complex and advanced

form of natural intelligence known. Intelligence indirectly created and developed through human intelligence is called artificial intelligence or machine intelligence. Artificial intelligence originates from natural intelligence, particularly human intelligence. Therefore, the primary task of AI research is to understand natural intelligence, create intelligent machines and applications, and enhance and serve the development of human intelligence [1]. The concept of artificial intelligence was first formally proposed by John McCarthy in 1955 at the Dartmouth Conference. Artificial Intelligence (AI), according to its English meaning, can be understood as "man-made intelligence."

Artificial intelligence is divided into three stages: weak AI, strong AI, and super AI. Weak AI utilizes intelligent technology to improve certain techniques needed for economic and social development. Strong AI is a stage very close to human intelligence, which is generally believed will not be truly realized until the mid-21st century. Super AI is an ultra-intelligent system that surpasses human intelligence, created after breakthrough developments in brain science and brain-like intelligence. Current AI development remains in the weak AI stage, with research fields primarily focusing on speech recognition, image recognition, natural language processing, intelligent robots, expert systems, and autonomous driving. The extended technological fields have already involved multiple aspects of human society [2].

3. Overview of Cyber Violence

This article examined the disciplinary research status of this concept through China' s largest literature database—China National Knowledge Infrastructure (CNKI)—which primarily involves multiple disciplines including journalism and communication, political science, law, and sociology, and has initially formed a scale (see Figure 1 [Figure 1: see original paper]). Many scholars have provided conceptual descriptions of cyber violence. For instance, Jiang Fangbing (2011) defined cyber violence as "a series of online deviant behaviors where network technology risks and offline social risks overlap through the interactive actions of network behavior subjects, potentially causing damage to the personality rights such as reputation and privacy of the parties involved" [3]. Chen Daibo (2013) considered cyber violence as "the general term for network aggressive behaviors implemented by netizens against parties or organizations to create psychological pressure and force them to yield; because this behavior has obvious coercive characteristics similar to violence in reality, it is called cyber violence" [4]. Lu Fang (2010) defined cyber violence as "a form of online behavioral deviance where network behavior subjects, as perpetrators, cause substantive harm to victims through concealed, coercive, extreme, and invasive network behaviors" [5]. Each of these scholars provides a conceptual definition based on emphasizing the nature and characteristics of cyber violence. Accordingly, this article summarizes cyber violence as: the general term for violent-like behaviors where network behavior subjects cause substantive psychological harm to parties through extreme and coercive means via online social channels.

The causes of cyber violence stem from multiple factors. First, the openness and anonymity of the Internet lead network behavior subjects (including individuals and commercial groups) to ignore objective facts in original social events. To increase attention or for commercial profit, they deliberately add subjective distorted opinions to create contradictions. After multiple rounds of forwarding and distortion, readers and viewers ultimately receive information completely inconsistent with objective facts. This network information distortion can anger some extremists and trigger cyber violence behavior.

Second, the youthful age of netizens and the generally low education level of the overall netizen population are also important factors. According to statistics from the "2021 National Minors Internet Usage Research Report" released by the China Internet Network Information Center (CNNIC): "In 2021, minor netizens reached 191 million, with internet penetration among minors reaching 96.8%, an increase of 1.9 percentage points from 2020 (94.9%)" . The 50th "Statistical Report on China' s Internet Development Status" shows: "As of June 2022, China' s netizen population reached 1.051 billion, with 19.19 million new netizens added since December 2021, and internet penetration reached 74.4%" . Data on netizen educational structure in the 47th "Statistical Report on China' s Internet Development Status" indicates: "Netizens with primary school, junior high school, and high school/technical secondary school/technical school education account for 19.3%, 40.3%, and 20.6% respectively; netizens with college, undergraduate, and above education account for only 10.5% and 9.3% respectively" . Through the above data, we can find that China' s netizens are mainly groups with medium education levels, and netizens are not only youthful but also generally have low education levels.

Third, with the explosive growth of the Internet digital economy in China, various technology companies based on Internet development have made further progress. Companies such as Tencent, Baidu, Sina, Youku, and ByteDance have all developed corresponding social apps, video software, and forums. The transition from the Web 1.0 era to the Web 2.0 era has provided network behavior subjects with dynamic, participatory, readable-and-writable, and convenient interactive methods, but it has also created a breeding ground for cyber violence [6]. Since cyberbullies themselves are also "customers" of such social platforms, driven by profit demands, these platforms tend to neglect self-regulatory mechanism construction, ultimately fostering the unhealthy trend of cyber violence behavior.

Finally, the lack of legal system construction for cyber violence is also a factor that cannot be ignored. Law is the basic norm for citizens' behavior, and targeted legislation is inevitably an effective means to control cyber violence. However, China still classifies cyber violence-related behaviors under the category of civil torts and has not yet formed special legislation on cyber violence. Once cyber violence occurs, victims can currently only rely on certain regulations and provisions to protect their rights. Additionally, due to multiple factors such as Internet anonymity, the diversity of cyber violence behavior subjects,

and the difficulty in estimating the degree of harm to victims, victims often fall into a dilemma of rights protection because of difficult evidence collection.

4. Feasibility of Precision Governance of Cyber Violence Through Artificial Intelligence

4.1 Urgent Need for Cyber Violence Governance

4.1.1 Purification of the Media Environment The "spiral of silence" theory posits that everyone is born afraid of isolation. Therefore, when facing controversial topics, they attempt to judge whether their opinions belong to the majority and whether public opinion will change in a direction that agrees with their views. Once they feel their opinions are in the minority, under the psychological effect of "fear of being isolated," they tend to remain silent on the topic. Consequently, the side with dominant opinion becomes more dominant, while the side with disadvantaged opinion becomes infinitely silent, forming a "spiral" communication process [7]. The Internet is an open world where anyone can express their views and opinions, but cyber violence accelerates the development of the "spiral of silence," hindering opportunities for minorities to express their views and destroying an open, inclusive, and free online media environment.

Moreover, the media plays a supervisory role in social governance, but cyber violence causes "supervision" to exceed "boundaries," causing illegal individuals, institutions, and commercial groups to turn supervision into "public opinion building"for profit. This "public opinion building,"combined with "clout-chasing" behavior in the self-media era, transforms benign media supervision into online behavior of "starting with a single image while fabricating the entire narrative," which not only harms public interests but also seriously reduces the credibility of mass media. The deterioration of the media environment caused by cyber violence urgently needs to be addressed.

4.1.2 Protection of Public Rights and Interests Cyber violence seriously damages public interests. The insults, rumors, verbal abuse, and other improper online behaviors by cyber violence behavior subjects can cause trauma to parties and their families, affecting the normal conduct of life. In recent years, many celebrities, scholars, and ordinary people have suffered from depression due to cyber violence and have even chosen to commit suicide. For instance, the Liu Xuezhou incident , the Wang Leehom divorce incident , and the TV drama cat abuse incident . At the same time, cyber violence distorts public moral values. The most common example of cyber violence events is moral judgment of parties based on personal moral standards, showing 偏激 and one-sided moral judgment. Due to Internet openness, anonymity, and corresponding legal deficiencies, this kind of online behavior easily exceeds normal boundaries. After obtaining the cathartic pleasure of cynical judgment, cyber violence behavior subjects cause spiritual devastation to parties without bearing any responsibility, and can even

demonstrate moral nobility and greatness. This unscrupulous online behavior mode distorts the public' s moral judgment standards and public moral bottom line [8].

4.1.3 Stable Development of Enterprises and the Nation Cyber violence hinders the healthy development of the Internet industry. As a "tumor" derived from the rapid development of the Internet era, cyber violence not only seriously affects the public' s online experience but also brings corresponding losses to network operators and social platforms, such as damaged brand image, loss of core users, and even 可能导致 business investment withdrawal, bankruptcy, and closure. Furthermore, cyber violence affects social stability and national security. Cyber violence behavior subjects use network platforms to disseminate aggressive and inflammatory speech, pictures, and videos related to sensitive topics. After diffusion and fermentation on the Internet, they appear in all corners of the network. This unlimited expandability and uncontrollability ultimately form a phenomenon of "one-sided online public opinion." Compared with traditional media public opinion, online public opinion has characteristics of multiple complexity, instant interactivity, and concealment. If the online public opinion formed by cyber violence involves topics such as "democracy" and "human rights," it can easily be exploited by Western or hostile forces to plan or incite reactionary activities, attack China' s political system, and thus affect China' s social stability and national security [9].

4.2 High Public Acceptance of Artificial Intelligence

Existing AI applications have had a tremendous impact on people' s lifestyles, learning methods, and entertainment methods. They not only greatly improve the efficiency of social resource utilization but also make production methods more humane. According to a Deloitte public opinion survey, "68% of Chinese users hold a positive attitude toward AI technology, believing that AI technology will have positive effects on social development, education, medical standards, environmental protection, and social fairness" . Using AI technology to solve cyber violence problems has become an increasingly hot topic in academia and society, and strong governance demand has prompted the further implementation and development of AI governance of cyber violence.

4.3 Strong National Policy Support

To promote the development of AI technology, the National Development and Reform Commission, the Ministry of Science and Technology, the Ministry of Industry and Information Technology, and the Cyberspace Administration of China jointly formulated the "Internet Plus Artificial Intelligence Three-Year Action Implementation Plan" (2016). The State Council issued the "Notice on the New Generation Artificial Intelligence Development Plan" (2017). The Ministry of Industry and Information Technology issued the "Three-Year Action Plan for Promoting the Development of the New Generation Artificial Intelli-

gence Industry (2018-2020)" (2017). The 2019 government work report further upgraded AI to "Intelligence Plus." National policy guidance and encouragement in the AI field have promoted China' s positive transformation from the "Internet Plus" era to the "Artificial Intelligence Plus" era.

On the other hand, to better prevent and control cyber violence problems and effectively protect netizens' legitimate rights and interests, relevant departments led by the Cyberspace Administration of China have actively carried out special actions against cyber violence and to improve the online environment, such as the "Clear and Bright Cyber Violence Special Governance Action" deployed and launched in April 2022 targeting 18 major network platforms, and the China Internet Civilization Conference held in August 2022.

4.4 Rapid Development of AI Supporting Technologies

"Data," "computing power," and "algorithms" are the three most important elements of artificial intelligence. They mutually promote and support each other, ultimately facilitating the application and value creation of AI technology. Since the 19th National Congress of the Communist Party of China, the construction of Digital China has achieved remarkable results, basically building the world' s largest and most technologically advanced network infrastructure. "By the end of 2021, China had built 1.425 million 5G base stations, accounting for more than 60% of the global total, with 355 million 5G users" . The improvement of network infrastructure has promoted the rapid development of China' s Internet industry. Cyber violence originates from the Internet, and each user interaction process in the Internet leaves backend traces. As interaction records continuously accumulate in backend systems, they ultimately form massive databases. The data stored in these databases provides prerequisites for AI training, and AI after skill acquisition can obtain corresponding intelligent models based on data from different scenarios. Computing power, namely the ability of computers to process data, is also the decisive factor for breakthrough AI development. According to data released at the 2022 China Computing Power Conference, "By the end of June 2022, China' s total scale of data center racks in use exceeded 5.9 million standard racks, with about 20 million servers and total computing power scale exceeding 150 EFlops. Meanwhile, the computing power industry chain continues to improve, including computing power infrastructure, computing power platforms, and computing power services. An internationally competitive computing power industry ecosystem has initially formed, with a batch of demonstrative computing power platforms, new types of data centers, and industrial bases being established" . The vigorous development of algorithm models, data technology, and the computing power industry has also laid a solid foundation for AI technology development, with companies and brands involving AI technology emerging, such as Baidu AI, Huawei, SenseTime, and Alibaba.

4.5 Practical Applications of Precision Governance of Cyber Violence

4.5.1 Sensitive Words, Sensitive Audio, and Sensitive Images and Videos For cyber violence behaviors expressed in the form of natural language descriptions, the most commonly used AI technology currently is "sensitive word filtering." By establishing a "sensitive word text description database" as the basis for detecting natural language descriptions, intelligent identification and filtering through the backend can directly shield, replace, or prevent the publication of filtered sensitive words, reducing the text dissemination of cyber violence sensitive words, thereby reducing language attacks and lowering language violence harm [10].

Since audio data is mainly in the form of waveform signals, for cyber violence behaviors expressed in audio form, waveform features are primarily used for extraction, identification, and classification of cyber violence characteristics. For cyber violence behaviors expressed in images and videos, it is generally necessary to extract image and video features. Using multi-classification algorithm models to identify different categories of cyber violence behaviors based on features such as motion, trajectory, and posture of targets in images and videos, and then locating images and videos with relevant semantic features, finally achieving the filtering and deletion of images and videos containing cyber violence such as blood and gore, terrorist organizations, firearms, pornography, spam marketing advertisements, watermarks, etc. Currently, major domestic network platforms have applications targeting sensitive words, images, videos, and audio, such as Sina Weibo, WeChat, and Douyin. Additionally, professional third-party content review services such as Baidu AI Intelligent Cloud, NetEase Yidun, and Shumei are becoming increasingly mature (see Figure 2 [Figure 2: see original paper]).

The AI "language reminder" and "user autonomous blocking" functions utilize AI technology. Before violent comments are published, the backend automatically sends reminders to commenters to change their language to maintain a friendly social environment. If a user' s personal account frequently receives abusive, attacking, or violent comments from one or more people, the user can set up a "blocking" function to prevent the continued publication of such comments. Both functions can help prevent cyber violence. Currently, many social platforms at home and abroad have launched such AI services. In 2019, the foreign image and video sharing social platform Instagram launched this kind of friendly speech reminder function. In February 2022, the image and video sharing social platform Douyin also became the first domestic platform to launch the "post warning" function. Additionally, Douyin added a new function to its cyber violence prevention system, namely the "Mood Warm Baby" platform assistant. After users repeatedly violate rules by publishing comments and private messages, AI automatically triggers the "Mood Warm Baby" to guide users to seek psychological assistance, medical treatment, and emergency help through online communication (see Figure 3 [Figure 3: see original paper]). Similar examples include the "Wali" robot assistant on the Zhihu platform.

5.1 Transition to a Contracted Government

Government size refers to an organic whole composed of various elements such as government institutions, functions, and personnel. Government size is not the smaller the better, nor the larger the better. Too small a government scale can easily lead to government failure and market failure, while too large a government scale can lead to institutional bloating and financial burden, which is not conducive to improving citizens' sense of happiness and government credibility. In emphasizing government governance capabilities, building a moderately sized government is also the basic orientation of contemporary public governance model transformation. AI development provides more efficient and convenient methods and means for processing massive public governance data. Through machine autonomous learning and precision algorithm models, AI can more scientifically organize and analyze massive data without interference from human subjective factors, thereby providing decision support for better governance solutions. Taking cyber violence governance as an example, AI can not only liberate governance subjects from simple and repetitive sensitive word filtering and deletion labor, reducing labor costs, but also help promote the flattening and networking of governance processes [11].

5.2 Transition to Precision Governance

Based on the practical applications of AI governance of cyber violence discussed above, with the support of big data technology, AI can accurately understand netizens' emotional changes and personal preferences after autonomous learning and creation. It can establish complete backend data files for each netizen and provide corresponding services and meet netizens' needs in a timely manner, providing a good prevention mechanism for cyber violence governance. In other areas of public governance, AI can still provide technical support for the precision and high efficiency of public services and governance.

5.3 Accelerating the Transformation of Public Governance Collaboration Methods

Public governance requires collaborative cooperation horizontally within public departments. At the same time, joint cooperation with external social organizations is also crucial. Governing cyber violence is not only the responsibility of Internet platforms; relevant public departments such as the Cyberspace Administration of China also bear regulatory responsibilities. Only on the basis of internal and external collaborative cooperation can the online environment be fundamentally improved and purified. Using AI' s automation and intelligent functions to obtain various social information needed for public governance can solve cooperation difficulties caused by information asymmetry between central and local governments, between localities, and between government and society. Building integrated and intensive intelligent centers can strengthen horizontal and vertical departmental cooperation in solving the "data island" problem.

6. Conclusion

From clarifying the concept, causes, and harms of cyber violence, combined with an overview of artificial intelligence, this article has analyzed the feasibility of precision governance of cyber violence through AI and its innovation in public governance. The application of AI technology can not only suppress the initial occurrence of cyber violence and precisely target cyber violence behaviors but also strengthen the prevention of cyber violence while guiding the benign and sustainable development of the Internet. Of course, cyber violence is a comprehensive social problem. Relying solely on AI technology to solve cyber violence-related problems is far from sufficient. In addition, it is necessary to continuously improve laws and regulations and formulate stricter regulatory measures to fundamentally solve the cyber violence problem. Although China' s development in AI technology is advancing rapidly, there is still a long way to go regarding AI governance of cyber violence.

Data source: August 23, 2022, using China National Knowledge Infrastructure (CNKI) journal database as the source. The retrieval process was as follows: select literature source as "Chinese General Database," conduct subject search using "网络暴力" (cyber violence), then tabulate the literature data volume in the "discipline" column.

China Internet Network Information Center, "2021 National Minors Internet Usage Research Report," http://www.cnnic.cn/n4/2022/1201/c116-10690.html, 2022-12-01/2022-12-05.

China Internet Network Information Center, 50th "Statistical Report on China's Internet Development Status,"http://www.cnnic.cn/n4/2022/0914/c88-10226.html, 2022-09-14/2022-12-05.

China Internet Network Information Center, 47th "Statistical Report on China' s Internet Development Status," http://www.cac.gov.cn/2021-02/03/c_{1613923423079314}.htm, 2021-02-03/2022-12-05.

Liu Xuezhou Incident—Liu Xuezhou, who was sold by his biological parents as a child, found his biological parents in 2021. As he was still a minor and economically dependent, with his adoptive parents deceased and homeless, he proposed living with his biological parents or having them provide housing, which they refused. Subsequently, online attacks accused Liu Xuezhou of being greedy, criticizing his appearance and voice. Ultimately, in January 2022, Liu Xuezhou committed suicide by taking medication on a beach in Sanya, unable to bear the pressure of cyber violence.

Wang Leehom Divorce Incident—In mid-December 2021, Li Jinglei published a blog post about the inside story of her divorce with Wang Leehom, initially gaining netizen support. However, a subsequent vague post mentioning a female singer surnamed Xu, combined with Wang Leehom' s delayed response, led to

the singer being cyberbullied, losing advertising contracts and income. Later, a writer surnamed Chen was also heavily cyberbullied for questioning Li Jinglei.

TV Drama Cat Abuse Incident—In a TV drama directed by someone surnamed Yu, a scene showing a cat dying from poisoning was suspected by netizens of being real. The production crew and director were cyberbullied despite multiple clarifications. The situation escalated, with over 60,000 people giving the drama low ratings, resulting in poor viewership. After the crew reported to police, three instigators of the cyber violence were arrested.

Deloitte "Artificial Intelligence Industry White Paper,"https://www2.deloitte.com/cn/zh/pages/innovation/ar
ai-industry-

"By the end of 2021, China had built 1.425 million 5G base stations, accounting for more than 60% of the global total, with 355 million 5G users."

According to the 2022 China Computing Power Conference, "By the end of June 2022, China' s total scale of data center racks in use exceeded 5.9 million standard racks, with about 20 million servers and total computing power scale exceeding 150 EFlops."

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.