RedFame

Studies in Media and Communication
Vol. 13, No. 3; September 2025
ISSN: 2325-8071 E-ISSN: 2325-808X
Published by Redfame Publishing
URL: http://smc.redfame.com

Artificial Intelligence, Social Media and Information Disorder in China's Information Ecosystem

Lishen Zhuang[1], Normahfuzah Ahmad[1]

[1]School of Multimedia Technology and Communication, Universiti Utara Malaysia

Correspondence: Lishen Zhuang, School of Multimedia Technology and Communication, Universiti Utara Malaysia, 06010 Sintok, Kedah Darul Aman, Malaysia.

Received: May 12, 2025 Accepted: June 23, 2025 Online Published: June 27, 2025

doi:10.11114/smc.v13i3.7792 URL: https://doi.org/10.11114/smc.v13i3.7792

Abstract

Artificial Intelligence (AI) technologies have played a significant role in the production and spread of information. AI technologies along with social media brought the public huge amount of information, which greatly improves the efficiency of information acquisition among societies. However, as AI technologies continue to advance, new forms of misinformation and disinformation emerge and disrupt the information ecosystem, resulting in the rise of various forms of fake news and information disorder. This poses serious threat to the authenticity and credibility of information and undermines healthy development of journalism in light of changing technologies. Given the limitations of the conceptual definition of 'fake news', this study is guided by Claire Wardle's broader and deeper framework of 'information disorder' and the Gatekeeper Theory. It adopts the qualitative method of in-depth interviews for data collection and utilizes thematic analyses to interpret its findings. The findings from this study expand ongoing literature on information disorder particularly within the fields of AI, social media and within the regional context of China.

Keywords: artificial intelligence, social media, information disorder, algorithms, gatekeeping theory, deepfake technology

1. Introduction

AI has been developing rapidly in recent decades as a highly visible and promising emerging technology (Gao et al., 2020). In terms of practical application, the introduction of AI technology into the field of news communication is a very wise choice, and its use has reshaped the entire news communication ecosystem. With the development of technology and the construction of various social networking platforms, the public has been able to move from an era of information scarcity to an era of information explosion, and the increase in the amount of information available has enriched the public's access to information and provided them with the ability to make better decisions for their lives and work. However, the digitalisation of news media and new affordances brought forth by digital technology have challenged the journalistic field (Ahmad, 2023). With the rise of AI technologies, the integration of AI into social media and news ecosystems has given rise to new forms of information disorder, which includes misinformation, disinformation, and malinformation (Wardle & Derakhshan, 2017). Information disorder occurs within the creation, production, and distribution of information, and the digital information system is increasingly characterised by its ecological nature (Chang, 2022). The opaqueness of information production is the key drawback presented by AI, where news sources and production processes are not accessible to the public (Hu & Cui, 2018). Moreover, social media has become nowadays a powerful source for fake news dissemination (Sharma et al., 2019). The use of algorithms in the dissemination process is not transparent and there is "virtual mining" of algorithm loopholes which puts users in an "information cocoon". The more audiences rely on the recommendations based on AI, the more likely the public is to be trapped in a filter bubble, which leads to the inability to access information that challenges or expands our world view (Pariser, 2011). The use of data sources needs to be regulated and faces the risks brought by new technologies such as big data (Wang, 2019). In the face of the current challenges, media professionals should be careful in choosing sources on social media, should not regard timeliness as more important than truthfulness, and should emphasize investigation and research reporting and truthfulness and pragmatism, or else they will be very easy to fall into a crisis in the risk of new technology (Xie, 2019).

In China, AI-generated content and algorithmic recommendation systems have dramatically lowered the cost of misinformation production, allowing users to quickly and easily generate highly realistic fake news (Lang, 2019). Prior to the emergence of deepfake technology, fake news was mostly presented in text, but as AI technologies have gradually infiltrated into short-form video production, applications such as AI anchors, ZAO and Deepfake with deep learning skills have made fake news appear spoofed and falsified (Zhai & Zou, 2020). The convergence of media and cross-platform communication further accelerates the spread of misinformation by amplifying the reach of fake news across different platforms (Livingstone, 2019). Social media platforms, in particular have become hotbeds for disinformation, where AI-driven recommendation mechanisms fuel the rapid spread of misleading content, often without proper verification or oversight (Adeline and Ahmad, 2022). As user-generated content continues to dominate digital spaces, the ability to distinguish authentic news from fabricated content becomes increasingly challenging, which undermines the public's trust in media institutions and the news ecosystem as a whole.

Artificial intelligence has gradually become the mainstay of news production and dissemination. When the traditional act of gatekeeping by news editors is replaced by a set of algorithms written by programmers, it leads to poor quality and error-ridden news content flourishing on the internet (Li, 2018). The weakening role of traditional media gatekeepers and the rise of digital platforms as new gatekeepers have shifted the power dynamics in the dissemination of news and information. Traditional editorial oversight, which once regulated and filtered content, is now largely replaced by algorithmic systems that prioritise engagement over truthfulness. As a result, fake news and misleading information often reach wider audiences more quickly, leaving the public vulnerable to disinformation (Li, 2018). AI technologies have contributed to the growing complexity of this disorder, making it more difficult to regulate and address. These developments have weakened traditional media gatekeeping mechanisms**,** distorted public perception**,** and threatened social trust in information sources.

The scope of the study centres on the Chinese digital media landscape, where AI technologies have played a transformative role in the production and spread of information. As AI technologies continue to advance, new forms of misinformation and disinformation are expected to emerge, further complicating efforts to combat information disorder. This study is based on the information disorder conceptual framework by Wardle and Derakhshan (2017) and the Gatekeeper Theory proposed to address the following research question:

What are the characteristics of information disorder popularised by artificial intelligence in China?

2. Literature Review

As the digital media ecosystem evolves, particularly with the advent of AI technologies, information disorder has become more complex and harder to regulate. Advances in AI technologies, along with the proliferation of social media increased the likelihood of information disorder spreading in societies. However, news organisations are continuously challenged to conduct holistic and effective verification with the rise of user-generated content from their audiences (Ahmad, 2023). Various studies have focused on different aspects of information disorder, particularly in relation to social media and algorithmic biases, offering valuable insights into its characteristics and consequences. Drawing on Wardle and Derakhshan's (2017) conceptual framework of information disorder and the gatekeeping theory, this study examines the rapid proliferation of information disorder that is frequently attributed to social media platforms in China, particularly through the increased use of user-generated content and AI technologies in media production.

2.1 The Conceptual Framework of Information Disorder

In September 2017, commissioned by the European Commission, the Shorenstein Center at Harvard University published the report *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*, in which Harvard Professor Claire Wardle, as the first author, formally proposed a new framework centered on the concept of "Information Disorder" (Er, 2021). Damasceno (2021) talked the term "fake news" is not used in his research because it does not fully capture the complexity of the problem. Instead, Wardle and Derakhshan's (2017) concept of information disorder is used, which includes the intentional and unintentional dissemination of misleading, false, or harmful news in the networked public spheres. "Information disorder" is not a new combination of words, but Wardle made a more academic and rigorous examination of its content from the level of methodology, and enriched its connotation (Er, 2021). Wardle (2018) pointed out that information disorder is not black and white, but has a flowing spectrum in order of harm from weak to strong: Satire or Parody, False connection, Misleading content, False content, Imposter content, Manipulated content, Fabricated content. Wardle and Derakhshan (2017) discuss these seven categories can be categorised into three types, based on truthfulness and intention to harm. Content that is false but not intended to harm is called misinformation. This can include satire, clickbait, or misleading quotes and images. Content that is false and intended to harm is considered disinformation and includes malicious lies, fabricated content, and manipulation campaigns. Finally, truthful information that is intended to harm is considered to be mal-information (Wardle & Derakhshan, 2017). This categorisation helps differentiate between the intentional and unintentional spread of false information and its potential

impact on public perception. According to Wardle (2018) the analysis of the elements and phases of information disorder helps to think about the characteristics and processes of information disorder communication. A message can be as simple as a tweet from an individual, or it can be an organised, long-term propaganda plan, and it can exist in the form of text, or it can be in the form of video, audio, and others. By using the conceptual framework of Information Disorder, including its three types and seven categories, this study deeply analyzes the characteristics of information disorder in China under AI technology and provides theoretical support for the formulation of effective coping strategies.

2.2 Fake News Escalation With Deepfake Technology

Deepfake is a combination of "deep learning" and "fake" as a form of "deep synthesis" techniques (Zhang, 2020). As a new form of forgery technology, creators can use it to tamper with the faces, voices, and other information of the characters in the video, so as to create short videos that are difficult to recognise as real or fake (Zhai & Zou, 2020). Deepfakes feature a cutting-edge technique that allows computers to create realistic videos of any person saying or doing anything (Damasceno, 2021). The representative application of "deepfake" is AI face-swapping, which is mostly used in the field of pornography, but now "deepfake" has begun to venture into other fields such as politics, advertising, science, and medicine. The developers of "deepfake" said that the technology is rapidly developing, and there is no limit to the objects they can imitate, which means that everyone is a potential target, and the future of deepfake bring a series of risks. Similarly, deepfake technology could pose a threat to journalistic authenticity (Xu, 2021). The Chinses cyberspace authority warned that "deepfake technologies could disrupt social order and violate people's interests, creating political risks and bringing a negative impact to national security and social stability" (Al Jazeera, 2019). In the era of AI, fake news has been transmuted under the influence of emerging technologies such as Deepfake, and the form has shifted from text to audio and video, which not only condones the generation of more fake news to a certain extent, but also spreads fake news in the form of short videos in an intelligent way, which is more infectious and confusing, and has a more tremendous negative impact on journalism. It also places higher demands on the audience's ability to discern information. In China, deepfakes are commonly known as Huanlian, which literally means "changing faces". Huanlian content, including face-swapped images and video reenactments, has been circulating in China since at least 2018, at first through amateur users experimenting with machine learning models and then through the popularisation of audiovisual synthesis technologies offered by digital platforms (De Seta, 2021). The launch of the ZAO app in 2019, which popularised the creation of face-swapped content and precipitated societal concerns around deepfakes, triggered a substantial regulatory effort from Chinese authorities (De Seta, 2021). "ZAO" is an infinite simplification of this technology; the user only needs a frontal photo to make an "AI face swap" to become a video subject.

2.3 The Gatekeeper Theory

Kurt Lewin introduced the concept of gatekeeping in 1947 (Thurman, 2015). Gatekeeping is a core concept of journalism and media research. News builds the basis upon which citizens form their political opinions. The media was retaining a traditional gatekeeping role, with journalists acting as message filters (Thurman, 2015). At the macro level, gatekeeping decisions shape what kind of political information is available within society. Gatekeeping is a concept about social power and the question of who controls the (political) information available within society. Since then, the concept has gradually evolved along with changes of media systems, providing gatekeeping scholars with the challenge to capture how changing media environments alter gatekeeping processes in journalism. Another key factor in the rise of information disorder is the weakening role of traditional media gatekeepers. In the past, media organisations through editorial processes, served as the gatekeepers of information, verifying facts before they were published. Li (2020) highlighted that the transition from professional journalism to user-driven content creation on social media has significantly diminished the effectiveness of editorial oversight. Instead of relying on professional journalists to filter and verify news, social media platforms have become the new gatekeepers of information, often without the necessary resources or protocols to ensure accuracy. In the online environment, news consumers are able to bypass traditional gatekeepers and seek information in meeting their interests, while ignoring the intermediary processors of news (Lee, 2012). Hence, the function of gatekeeping "has shifted from the decision about what should be produced to control of what materials get to consumers and of what material they become aware" (Hargittai, 2004). Many scholars have started to question whether gatekeeping can be a tenable theory in the decentralised, new media. Digitisation and the Internet were said to lead to the end or "death" of gatekeeping in journalism (e.g., Kovach & Rosenstiel, 1999; Lasica, 1996). However, this trend does not eliminate entirely the process and role of gatekeeping.

2.4 Artificial Intelligence and Cross-Media Convergence

The increasing prevalence of AI-driven news automation also plays a significant role in the proliferation of information disorder. According to Lang (2019), the development of AI technologies has reduced the cost and effort involved in producing and disseminating fake news, thus facilitating its spread. AI-based news generation systems can produce highly realistic news reports, which are often indistinguishable from genuine news articles (Lang, 2019). These automated

systems have been used to create news narratives that align with specific agendas or biases, further distorting public perception and reinforcing pre-existing biases (Chang, 2022). The cross-platform nature of modern media has also intensified the spread of misinformation in China. Livingstone (2019) noted that media convergence allows content to flow seamlessly across different platforms, further amplifying the spread of false narratives. For instance, a piece of disinformation initially shared on WeChat can quickly spread to other platforms like Weibo and Douyin, reaching a much larger audience in a shorter amount of time.

The characteristics of information disorder in China are deeply intertwined with the rapid advancements in AI technologies and the increasing reliance on social media platforms for news dissemination. The AI-powered creation of misleading content, the weakening role of traditional gatekeepers, and the amplification of false information through algorithmic recommendations have all contributed to the growing complexity of information disorder. For example, the recent popularisation of DeepSeek, a Chinese-based AI company has benefited many in China, but simultaneously, its adoption raises risks to the susceptibility of fake news created by AI (Reuters, 2025). As media convergence and cross-platform communication continue to evolve, the challenges associated with regulating misinformation are likely to intensify.

3. Methodology

This study used qualitative research methods to conduct semi-structured in-depth interviews with professionals in China's media industry. Qualitative methods explore the perspective and meaning of experiences, seek insight, and identify the social structures or processes that explain people's behavioral meaning (Mays & Pope, 2000). Informants were identified through recommendations from journalists and industry experts, and further correspondence was established with relevant staff members via email, Weibo, Zhihu, and other social media platforms. The selected informants included journalists, producers, and AI experts. It is argued that the number of participants selected is not a fixed value, but is based on the principle of data saturation (Creswell, 2013). Therefore, for this study, the final determination of 13 informants was based on the saturation of the data obtained during field work. The purpose of the interviews was to collect information disorder characteristics of the popularity of AI in China. Online and face-to-face interviews were planned according to the informants' availability to ensure convenience and accessibility for the informants. These informants were professionals who had direct experience, interaction or process. Where possible, gender, age, occupation, and educational background were chosen to ensure that the diversity of the sample was sufficient to be able to have a full perspective and study the data. Interview outlines were developed based on the research questions, and each interview lasted approximately 50 to 90 minutes, with the final duration of the interview ending when data saturation was reached.

In addition to primary data collected from interviews, secondary sources we obtained by analysing literature, reports and academic papers. Secondary sources were collected from materials published between 2010 and 2025, from which this study will gather literature and cutting-edge ideas related to AI, information disorder, fake news, and deepfake, to understand the current state of national and international research, and to seek out a variety of sources to learn about the theme. These secondary sources will help to contextualise and support the findings of the interviews. When collecting data from secondary sources, attention is paid to the availability, format and quality of the data to ensure the relevance of the information.

The data analysis was carried out using thematic analysis. The interview transcripts and secondary sources were coded to identify key themes. Thematic analysis allows for the identification of patterns and connections between the themes, providing a comprehensive understanding of the characteristics of information disorder. Ethical considerations are a priority throughout the study. Informed consent was obtained from all informants prior to interviews to ensure that they were fully aware of the purpose of the study and their right to confidentiality and anonymity.

4. Results

Findings suggest that China's information disorder in the era of AI is characterised by various aspects. These are elaborated next.

4.1 The Characteristics of Social Media

In the era of AI, social media relies on algorithmic recommendation technology to become the current traffic congregation in cyberspace; however, more advanced information distribution technology has not solved the information disorder but rather caused the negative impact to be more serious (Chen & Chen, 2023). Social media have greatly reduced the cost of production and the threshold for dissemination of disinformation. At the same time, algorithmic recommendation mechanisms and cross-platform interactions have further accelerated the speed and scope of the dissemination of information disorder, especially in the fragmented and short-term information environment, where the efficacy of the dissemination of information disorder has been amplified. The recommendation systems of social media platforms act as triggers and pushers of information disorder (Chen & Chen, 2023).

Firstly, social media is covert and the source is opaque. The hidden and opaque nature of social media exacerbates the uncertainty of where people get their information from, thus further expanding the potential for information disorder (Zhao, 2024). Informant 10 talked about the generalisation of the subject of fake news reports and the difficulty of identifying the source of information.

> *"Social media platforms allow users to post content anonymously or under a pseudonym, which makes it difficult to trace the source of information. This anonymity and invisibility makes it easy for malicious actors to spread misleading or false information, leading to information disorder" (Informant 1).*

Secondly, the lower cost of producing false information. Advances in digital technology and AI have made the production of false information cheaper than ever before. In the past, creating false information required a certain level of expertise and resources, but today, even ordinary users can quickly produce highly realistic content through free or low-cost software tools.

> *The proliferation of modern digital tools and software has made it possible for individuals to create high-quality multimedia content without the need for specialised skills. The fact that most of these tools are free or low-cost further lowers the threshold for the production of fake news. It makes it easier for false information to be produced and disseminated (Informant 4).*

Thirdly, enhanced media convergence, combined with cross-platform communication and fragmented communication modes drive the rapid proliferation of false information. Livingstone (2019) highlighted that the contemporary information ecosystem is characterised by deep media convergence and seamless cross-platform interactions, allowing disinformation to spread effortlessly across multiple channels. For example, a false narrative initially posted on Twitter can quickly circulate on Instagram, Facebook, or WhatsApp, reaching a broader audience. Additionally, the rise of short video platforms like DouYin and TikTok has further exacerbated information disorder. Their intuitive, short-form content and ease of sharing make them powerful vehicles for the rapid dissemination of false information, as seen during COVID-19 in China. Informant 8 talked about the content interoperability between platforms enhances the speed of rumour spreading.

> *In China, the social media ecosystem is highly developed, and the flow of information between multiple platforms is extremely strong. Due to social media algorithmic recommendations and simultaneous reprinting on multiple platforms, false news undergoes constant processing, distortion, and amplification in the process of dissemination, and the user groups of Chinese social media platforms (e.g., Weibo, WeChat, DouYin, KuaiShou, Xiaohongshu, and Bilibili) overlap, with users often spreading their content across different platforms (Informant 8).*

Fourthly, the role of gatekeepers has been weakened. In the digital age, this function has been significantly weakened by the rise of social media and user-generated content (UGC). Liu (2021) pointed out that the core mechanism of social media platforms relies on users to post content on their own, a model that improves the efficiency of information dissemination but lacks effective vetting of the authenticity of information, making it easier for false information to spread in the public sphere. For example, a large number of fake accounts and zombie websites use social media to publish inaccurate information, which not only disrupts public opinion, but also seriously erodes the information ecology. Informant 2 shared a similar view that the diluted role and weakened role of gatekeepers allowed online fake news to take advantage of the situation.

> *If online media journalists do not follow the normal news release process and ignore or filter out the gatekeeper role, online fake news will inevitably occur (Informant 2).*

The four core features reveal how social media in the era of AI has become an amplifier of information disorder. These structural features provide a breeding ground for the widespread penetration of deepfake technology. Through its ability to generate highly realistic content, deepfake technology has completely breached the boundary between truth and fiction, pushing information disorder to a new and more destructive stage. The following section will analyse another factor in the characteristics of information disorder.

4.2 Widespread Use of Deepfakes

"Deepfake" can synthesise pictures, videos, and voices according to human behavioural characteristics to complete face replacement (Zhang, 2020). The representative application of "deepfake" is AI face-swapping, which is mostly used in the field of pornography, but now "deepfake" has begun to enter other fields, such as politics, advertising, science and medicine (Xu, 2021). Deepfake technology enables the synthesis of fake videos, images, and other media content, posing a severe threat to the authenticity and credibility of news. Malicious actors exploit AI-driven image-generation algorithms to fabricate realistic fake news images, synthesise deceptive video scenarios, or mimic public figures' voices to disseminate false statements or misleading opinions.

In China, Deepfake technology began to invade and disrupt the normal order of communication, from the star face-swapping video produced by a Bilibili user in February 2019, to the face-swapping app 'ZAO', which was an overnight sensation in China's app market in September, but was quickly taken off the shelves (Zhai & Zou, 2020). The characteristics of 'face-swap video', such as difficulty in distinguishing the real from the fake, strong concealment, fast dissemination and wide dissemination, will also bring about serious effects, including the proliferation of false videos (Sun & Zhou, 2020). Consequently, Informant 5 clarified that deepfake, as a new characteristic of information disorder, is extremely disruptive to public perceptions.

> *The application of deepfake technology has greatly reduced the cost of producing and circulating false and malicious information, in which information about celebrities, women, and vulnerable groups is more likely to be tampered with or faked by AI-generated false content（Informant 5).*

There has been an exponential growth in 'fakes' on the US web, with 14,678 fake videos in July 2019, 96 percent of which were pornographic, and climbing to 49,081 in June 2000 (Liu, 2021). Informant 9 also indicated that the stealthy and disorientating nature of deepfake technology is also due to its low threshold and high popularity. As technology matures, ordinary users can also produce high-quality fake content through simple software tools.

> *In China, AI video generation technology is developing rapidly, and there are currently a number of AI video generation software on the market with a low threshold for soft use. This democratisation of the proliferation of technology has made it easier to disseminate false information, which not only poses a threat to the image of public figures, but has also been used to target groups of ordinary people (Informant 9).*

"Face-swapping video" method of forgery, not only undermines the authority of the dynamic video verification of the real but also will further enhance the difficulty for the public in the information dissemination activities to obtain the real information (Sun & Zhou, 2020).

> *The false information generated by deep synthesis technology is indistinguishable from the authentic version, which greatly disturbs public perception. It can also lead to a crisis of authenticity (Informant 13).*

Deepfake not only achieves realistic replacement of faces, voices and behaviours, but also shifts the production of disinformation from specialisation to massification through social media. Deepfake technology makes false information break through the traditional verification mechanism through high concealment and strong confusion. The public is plunged into a crisis of trust because it is difficult to recognise the authenticity. Deepfake is not only a link in the chain of information disorder, but its harm is further complicated by the algorithmic system. The next section will analyse another characteristic of information disorder.

4.3 Algorithmic Bias

The disorientation of false information is also closely related to algorithmic bias. Intelligent recommendation algorithms aim to deliver content accurately based on users' interests and behavioural preferences through data analysis and pattern prediction. Understanding consumers' interests enables social media to provide them with content that keeps them interested at all times, thus prolonging users' time on the social media platform (Adeline & Ahmad, 2022). However, such personalised recommendation mechanisms act as amplifiers in the dissemination of false information. Pariser (2011) introduced the concept of 'information cocoon' which describes how personalised algorithms reduce the diversity and transparency of information by exposing users only to information that is consistent with their existing preferences and beliefs. The information cocoon makes it difficult for users to access different viewpoints or evidence to refute disinformation, the role of traditional editors as gatekeepers has been weakened to a certain extent, and users have become precise 'targets' (Chen & Lv, 2022). Algorithmically recommended content often lacks vetting and gatekeeping of content and the quality of information cannot be guaranteed, and it is prone to contain fake news or extreme views (Wang & Li, 2018).

> *Algorithmic recommendations in social networks drown out moderate views but contribute to the spread of extreme views. Whether algorithms are able to distinguish between 'fact' and 'fiction', and whether they lead to the widespread dissemination of low-quality news and fake news are technical issues that still need to be resolved in terms of algorithmic recommendation (Informant 6).*

AI algorithms are biased, leading to unfair information recommendations, exacerbating the information cocoon effect, and affecting the public's right to know and to choose. Informant 3 said that the precise push of AI algorithms creates a cocoon of information and an echo chamber. Informant 5 talked about the application of AI technology in news production lacks the depth and perspective of human journalists, and algorithmic bias can lead to unfair news reporting.

> *Algorithmic bias can easily lead to information disorder, and algorithms may recommend news based on the biases in the training data. The information cocoon effect may be exacerbated, and users may only be exposed to information that agrees with their own views (Informant 11).*

Fake text, images, voice and video content that can be automatically generated by AI technology. Similarly, fake press releases can be generated using AI-generated technology and disseminated quickly through algorithmic recommendation mechanisms. Informant 1 mentioned that some self-publishing platforms can post dozens of messages a day, and it's hard to believe that it's done by humans. As well as Informant 9 also talked about how AI technology tends to enable the automated generation of fake news.

> *AI technology has led to an increase in the amount of automatically generated and disseminated fake news that may be difficult to identify by traditional methods. AI technology supports the spread of false information faster and wider, making it difficult to dispel rumours. After a false message has spread, it takes several days for the disinformation message to reach most users. An AI-generated fake video of a 'ocial event' on a short video platform received more than 10 million views in one day, and when it was clarified afterwards, the impact was irreversible (Informant 10).*

The perplexing nature of algorithmic recommendations is particularly reflected in its prioritisation of pushing false information. Vosoughi et al. (2018) found that false news spreads six times faster than real news and reaches a wider range of users by analysing Twitter data. This is mainly due to the fact that fake news is usually designed to be more eye-catching, easily triggering users to click and retweet, and these behaviours are recorded by algorithms and used to further optimise content recommendations. Informant 1 mentioned that the popularity of AI technology is prone to the formation of algorithmic recommendations, information cocoon and other new problems.

> *When in an information cocoon for a long period, algorithmic recommendation reinforces the original bias of the audience, and it is difficult to access diversified content and facts, their judgement ability may be weakened, leading to the further solidification of misperceptions, and ultimately falling into a state of total skepticism about the authenticity of the information (Informant 1).*

Algorithmic rules are made by people. People disseminate information in social media, but at the same time, it can be 'manufactured' by interest groups to manipulate public opinion. Algorithms are influencing social reality in a covert way, creating a 'technological unconscious'. Informant 5 also believed that algorithms are prone to manipulation of public opinion.

> *Algorithmic systems are even more ubiquitous in social media, where algorithms are used to guide users to receive information, optimise the ordering of popular trends, and collaboratively filter information, leading to significant changes in both the habits and cognitive ways in which audiences access information (Informant 5).*

In the era of AI and big data, algorithmic recommender systems have become the core driver of social media and information platforms, but while it enhance the user experience, they also raise serious data privacy protection issues. The large-scale collection, storage and utilisation of user data has led to increasing privacy leakage, data abuse and information security risks. Informant 1 mentioned that the gap between technological development and public awareness of privacy.

> *The collection of personal information on social media and how to balance and protect the privacy of users is something that needs to be discussed. China's science and technology are developing at such a fast pace that some areas are not yet ready. There exists one, a big gap between the development of science and the perceptions (regarding its societal impact, ethical implications, and personal privacy concerns) of people. Secondly, there is a big difference between the rules and legislation of our country and the comparison of some developed countries (Informant 1).*

With the development of algorithmic technology, people began to rely on algorithms to solve many problems, especially when our intuitive judgment is problematic is making a decision, algorithms become a lifesaver, which has led to very extreme algorithmic dependencyism. Informant 9 talks about how the logic hidden in the algorithms is constructing the rules of the world's operation, in which the rules of operation that internalise human logic are internalised into a procedure, which in turn comes back to influence humans.

> *Algorithms are decision rules generated through historical data, and they are also a mirror image of society and a reflection of the human world. Then, the good and bad, cultural differences, and conscious bias that humans have will be reflected in the algorithm. Therefore, the values of communication in the age of algorithms are still guided by humans. The values of algorithms depend on the data, and the data depend on humans themselves, so the scientific nature of the data as well as the value of the data is very important (Informant 9).*

In an age of intertwined AI and social media, the way information is disseminated has changed profoundly. Social media's algorithmic recommendation mechanisms amplify misinformation and accelerate the spread of unverified content. Deep faking techniques further enhance the deception of false information, making it harder to detect maliciously created fake news. Algorithmic bias reinforces the information cocoon, shaping audience perceptions and weakening their ability to distinguish between true and false information. The traditional media gatekeeper role has been greatly diminished. Social media and algorithms dictate the spread of information, shifting the gatekeeping function to users who often lack

verification skills. The concealment and proliferation of false information has accelerated significantly, and the fragmented distribution model of social media and the widespread use of deepfake and algorithmic techniques, along with the weakening of gatekeepers, have combined to exacerbate information disorder.

5. Discussion

The findings of this study highlight the significant role that social media characteristics, deepfake technology, and algorithmic bias play in fostering information disorder in the context of AI-driven media environments in China. The findings resonate with Wardle's framework, particularly in how social media platforms facilitate the rapid spread of misleading content while blurring the boundaries between different characteristics of information disorder.

The study found that social media's covert nature and the opacity of information sources significantly contribute to information disorder. The lack of transparency regarding where information originates increases doubts among users. This is added by the reduced cost of creating misinformation, making it easier for individuals to produce and spread fake news. These characteristics of social media platforms, combined with the media convergence and cross-platform dissemination of content, accelerate the spread of misinformation and contribute to the fragmentation of information. These findings are consistent with existing literature that stresses the interconnectedness of social media and misinformation.

Moreover, the application of deepfake technology has exacerbated the issue of fake content generation. From a gatekeeping perspective, traditional media institutions have historically played a role in filtering and verifying information before dissemination. However, in the AI-driven media landscape, the weakening of these traditional gatekeeping mechanisms allows manipulated content, such as deepfakes, to circulate more freely, challenging the role of journalists and fact-checkers. This phenomenon highlights the need for advanced technological solutions to combat deepfakes and ensure the credibility of information.

Finally, algorithmic bias identified as a critical factor in shaping the information ecosystem was found to reinforce existing beliefs and reinfore information cocoon among societies. In the digital age, algorithms have assumed the role of gatekeepers, selectively amplifying or suppressing information based on engagement metrics rather than journalistic standards. These findings highlight the evolving dynamics of information flow and the increasing challenges brought by technological challenges of AI in the media environment.

6. Conclusion

The threat of information disorder has become increasingly severe in today's digital society, particularly under the influence of AI and social media. The main objective of this study was to analyse how AI impacts the characteristics of information disorder. The study found that the opacity of social media and the lack of transparency in information sources have made the spread of false information much easier, while also reducing the verifiability of information. The fragmentation of information and cross-platform dissemination further exacerbate this issue, with fake news rapidly spreading across various platforms, leading to a crisis of public trust in information.

Additionally, the widespread use of AI technologies has made the production of fake information cheaper and more efficient, further intensifying the problem of information disorder. Deepfake technology makes fabricated content nearly indistinguishable from real content. The shift from human editorial control to algorithm-driven content curation has reshaped the traditional media landscape. Algorithms now function as digital gatekeepers, selecting and amplifying content based on user engagement rather than journalistic integrity. However, algorithm-driven content has its disadvantages since it reinforces echo chambers and intensifies information disorder distribution in societies.

An effective method to address these challenges is to build a media ecosystem centred on standardised journalistic ethics, ensuring the accuracy and fairness of news reporting. Therefore, all forms of media must adhere to ethical principles in journalism and emphasize accuracy and reliability to ensure their sustainability in the future. This study contributes new perspectives to the existing literature on information disorder particularly in the context of AI and social media in China. The study highlights the urgency of strengthening algorithm regulation and deepfake detection technologies towards a more healthy and truthful information ecosystem in light of challenges brought by AI and its technological changes in societies.

Acknowledgments

Thank you to all the informants who participated in this study.

Authors contributions

Lishen Zhuang was responsible for drafting and revising the manuscript, research design and data collection, while Dr Normahfuzah Ahmad was responsible for reviewing and editing it. All authors read and approved the final manuscript.

Funding

This research received no external funding.

Competing interests

The authors declare no conflict of interests in any forms.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Redfame Publishing.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

References

A Ga Er. (2021). From Fake News to Information Disorder: The Challenge of MIL Education in the New Vision. *New Century Library*. *21*(02): 5-10.

Adeline, H. D., & Ahmad, N. (2022). Why do people fall for fake news? Patterns in news consumption and decision-making. *Journal of Media and Information Warfare, 15*(3), 27-38.

Ahmad, N. (2023). Journalistic verification practices from the BBC World News and Al Jazeera English. *Howard Journal of Communications,* 1-14. https://doi.org/10.1080/10646175.2023.2233096

Al Jazeera. (2019, November 30). *China criminalizes 'fake news' created with AI, bots*. Al Jazeera News. https://www.aljazeera.com/news/2019/11/30/china-criminalises-fake-news-created-with-ai-bots

Chang, H. (2022). A brief discussion on the application of artificial intelligence combined with journalistic ethics. *Shanxi Economic Daily*, *23*(7), 1-2.

Chen, C. F., & Lv, Y. X. (2022). *Research on Algorithmic Ethics: Perspectives, Frameworks and Principles. Inner Mongolia Social Sciences, 43*(03), 163-170 + 213. https://doi.org/10.14137/j.cnki.issn1003-5281.2022.03.022

Chen, X. J., & Chen, L. J. (2023). Information Disorder Risks and Governance Paths on Social Media Platforms. *News World,* (10), 53-55. https://doi.org/10.1007/s15036-023-3302-3

Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among the five approaches (3rd ed.).* Thousand Oaks, CA: Sage.

Damasceno, C. S. (2021). Multiliteracies for combating information disorder and fostering civic dialogue. *Social Media + Society, 7*(1), 205630512098444. https://doi.org/10.1177/2056305120984444

De Seta, G. (2021). Huanlian, or changing faces: Deepfakes on Chinese digital media platforms. *Convergence, 27*(4), 935-953. https://doi.org/10.1177/13548565211030185

Gao,W., Xiao, Z. H., & Zhu, Y. L. (2020). The Security Problems of the DeepFake Technology: Opportunities and Challenges. *Journal of Information Security Research*, *20*(06), 634-644.

Hargittai, E. (2004). The changing online landscape: From free-for-all to commercial gatekeeping. In P. Day & D. Schuler (Eds.), Community practice in the network society: Local actions/global interaction (pp. 66-76). New York, NY: Routledge.

Hu, H., & Cui, Y. J. (2018). Building a "good partner" for the real world of news exploration of the real world of news based on AI and blockchain technology. *Journal of Qingdao University of Science and Technology (Social Sciences), 34*(04), 96-98.

Kovach, B., & Rosenstiel, T. (1999). *Warp speed: America in the age of mixed media.* New York, NY: Century Foundation Press.

Lang, J. H. (2019). 2018 Fake news characteristics and countermeasures study. *News Dissemination, 19*(7), 12-13.

Lasica, J. D. (1996). Net gain: Future of the news on the Internet. *American Journalism Review, 18*, 9-20.

Lee, S. H. (2012). The end of the traditional gatekeepers. *Journal of Communication, Culture & Technology, 12*, 1-24.

Li, X. S. (2018). *Changes in news production in the post-information era under artificial intelligence conditions* (Master's thesis, University of Chongqing Technology and Business).

Li, Y. (2020). *The influence of Artificial Intelligence on Political elections* (Master's Thesis, Jinan University).

Liu, H. M. (2021). Discussion on the Information Dissemination Characteristics and Governance Paths of Short Video Platforms. *Modern communication, 21*(05), 34-42.

Livingstone, S. (2018). Audiences in an age of datafication: critical questions for media research. *Television & New Media, 20*(2), 170-183. https://doi.org/10.1177/1527476418811118

Mays, N., & Pope, C. (2000). Qualitative research in health care: Assessing quality in qualitative research. *The BMJ*, *320*(7226), 50-52. https://doi.org/10.1136/bmj.320.7226.50

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you.* Penguin UK. https://doi.org/10.3139/9783446431164

Reuters. (2025, March 15). China to crack down on stock market fake news as AI spurs misinformation, says state media. https://money.usnews.com/investing/news/articles/2025-03-15/china-to-crack-down-on-stock-market-fake-news-as-ai-spurs-misinformation-says-state-media

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating Fake News: A survey on identification and mitigation techniques. *arXiv (Cornell University).* https://doi.org/10.1145/3305260

Sun, X. H., & Zhou, S. Y. (2020). "Ethical risks and responses arising from 'face-changing Videos' in Communication. *Youth Journalist*, (11), 22-23. https://doi.org/10.15997/j.cnki.qnjz.2020.11.011

Thurman, N. (2015). Journalism, gatekeeping and interactivity. In *Edward Elgar Publishing eBooks*. https://doi.org/10.4337/9781782548768.00031

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146-1151. https://doi.org/10.1126/science.aap9559

Wang, B. & Li, W. Z. (2018). How to Break Through the Cognitive Narrowing in News Filtering Bubble Algorithm Push Notifications and How to Avoid It. *Journal of Journalism and Writing,* (09), 20-26.

Wang, X. Z. (2019). *Research on News Production Based on Artificial Intelligence Technology*. (Master's thesis, Jiangxi University of Finance and Economics, China).

Wardle, C. (2018). The need for smarter definitions and practical, timely empirical research on information disorder. *Digital Journalism*, *6*(8), 951-963. https://doi.org/10.1080/21670811.2018.1502047

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking* (Vol. 27, pp. 1-107). Council of Europe.

Xie, G. Y. (2019). The impact of artificial intelligence technology on fake video news and coping strategies from the "ZAO" incident. *News tide, (*12), 10-11.

Xu, Y. P. (2021). Exploration of the path of news fact-checking in the context of "deepfake". *New media research. 21*(4), 08-17.

Zhai, H. L., & Zou, X. C. (2020). The evolution and governance of short-form video-based fake news with the support of Deepfake. *Today's Mass Media*. *20*(11), 34-36.

Zhang, T. (2020). Legal risk and regulation of deep forgery in the post-truth era. *E-Government. 20*(8), 91-100.

Zhao, L. W. (2024). Analysis of Information Disorder in the Era of Intelligent Media and Media Response Strategies. Journal of Reporting, *Writing and Editing,* (06), 48-50.