Proceedings of the 2nd International Conference on Social Psychology and Humanity Studies
DOI: 10.54254/2753-7048/50/20240802

Research on Algorithmic Gender Bias under the Paradigm of Machine Behavior Studies

Yuqing Liu[1,a,*]

[1]*Hunan Normal University, 36 Lushan Road, Changsha City, China*
a. 1274920351@qq.com
**corresponding author*

Abstract: In the field of communication studies, machine behavior specifically refers to information dissemination activities involving artificial intelligence technology. As algorithms increasingly become the primary force in information dissemination, their potential gender bias becomes increasingly apparent. This paper, based on three research scopes in machine behavior studies: individual behavior, collective behavior, and human-machine interaction behavior, examines the gender bias exhibited by artificial intelligence entities in algorithms at these three levels. At the individual behavior level, the tendency of algorithm development to simplify features overlooks the diversity present in female society. The inherited data bias and human bias make it difficult to avoid gender discrimination. At the collective behavior level, the creation of opinion leader-type social robots expands the subject of information fog, making the concealed “gender discrimination against women” more covert. The use of large-scale machine armies manipulates search engine results, leading to severe gender bias in search engine outputs. At the hybrid human-machine behavior level, artificial intelligence shapes female images to construct female cognitive thinking. Algorithms acquire human bias during interaction with users, and social robots amplify gender bias issues through mixed human-machine behavior.

Keywords: Machine Behavior Studies, Algorithm, Gender Bias

1. Introduction

In April 2019, Nature published an article entitled “Machine Behavior”, announcing the official birth of “machine behavior”, an emerging discipline spanning multiple research fields. The intention of Machine Behavior Studies is to scientifically explore intelligent machines, The core concept of “machine” does not solely refer to technology or physical machines but broadly encompasses all artificial objects and phenomena. In a broad sense, machine behavior represents the shaping activities of machines driven by artificial intelligence technology in society, such as automatic driving technology.[1] This paper, from the perspective of communication studies, defines “machine behavior” as information dissemination activities involving artificial intelligence technology, such as algorithmic recommendations and AIGC.[2] The birth of the machine behavior studies paradigm provides communication researchers with a new research perspective, including the issue of algorithmic bias introduced by artificial intelligence technology.

The advancement of technology continuously reshapes the social environment in which humans live. Automation decisions implemented through algorithms have been widely applied in various scenarios. Various algorithmic recommendations and models increasingly influence our cognition and decision-making, thereby gradually revealing the issue of algorithmic bias. Concerns arise that discrimination based on factors such as gender, race, age, and social status may spread to artificial intelligence and algorithmic technologies. This may replicate biases, injustices, and inequalities from reality into the technological domain, exacerbating social divisions and ideological splits. In fact, these concerns have become a reality to some extent. In response to gender bias issues in artificial intelligence research and applications, UNESCO released its first recommendation document, "I'd Blush If I Could," in 2019. The document explores the persistent gender gap in digital skills globally, emphasizing that the application of artificial intelligence can perpetuate harmful stereotypes against women.[3] It calls for new commitments and transformative actions to promote gender equality. In light of this crucial finding, a global dialogue is needed on the relationship between artificial intelligence and gender bias because AI methods clearly deviate from gender fairness, which could lead to constructive social changes. For example, Reuters reported in 2018 that Amazon's automatic recruitment system exhibited gender discrimination issues. The system, when screening resumes, would downgrade female applicants and elevate the rankings of male candidates.[4] This illustrates that various recruitment software, while eliminating gender bias against women, also facilitates employers in using gender-sensitive language when crafting more inclusive job postings.

2. Literature Review

The article "Machine Behavior" proposes three research scopes for the study of machine behavior: individual behavior, collective behavior, and human-machine interaction behavior. These scopes emphasize the research on algorithms themselves, the interactions between machines, and the interactions between machines and humans, respectively. [1]From a macro perspective, these three research scopes in machine behavior studies can be seen as an in-depth investigation into a specific species, exploring interactions among the species and their interactions with a broader environment. This provides a comprehensive research structure for our study in the field of communication on the behavior patterns of artificial intelligence and their impact on society. However, to date, research in the field of communication studies on machine behavior has largely focused on social robots.

2.1. Individual Behavior Level

Social robots, defined as "virtual AI entities that play the role of human identity, possess varying degrees of personality attributes, and interact with humans in social networks,"have now widely infiltrated social media composed of diverse circles and subjects.[5] [6] Social robots disrupt the traditional opinion formation process and, in their participation in the process of opinion exchange, reconstruct the information access rights of information entities. [7]

Firstly, at the individual behavior level, machines, particularly social bots, engage in shaping the construction of public opinion in social networks as highly personified opinion leaders. They concentrate discourse resources in information dissemination, thereby enhancing the influence of the positions they represent. Typically, in the environment of social media, opinion leaders can attract more public attention by conveying issues, simultaneously arousing the interest of media, stakeholders, and policymakers, thereby promoting public participation.[8] Opinion leader-type social bots are commonly observed in propaganda activities during conflicts. During the Syrian war, for instance, a Twitter social bot account named "@sahouraxo," disguising itself as a geopolitical commentator independent of Lebanon, with 125,000 followers, garnered widespread attention. The influence of this account almost reached a level consistent with some BBC journalists continuously reporting on the Syrian war. [9]

Furthermore, social robots can leverage the prevailing emotions on social media to either amplify or diminish the "backfire effect" and manipulate false trends, thereby altering public opinion.[10] Social robots are often tasked with presenting a particular opinion in a attention-grabbing manner, specifically for "refutation." In this process, individuals who disagree with the stance of the social robot are more likely to take explicit actions, such as commenting or liking, to express their opposition to the "refuted opinion." Therefore, to guide social network discourse in a direction aligned with societal goals, social robots may intentionally refute consensus views to persuade the audience. Conversely, to guide discourse in a direction contrary to societal goals, they may adopt a strategy of initially conforming to public opinion before expressing dissent, aiming to persuade the audience. As an example of the latter case, a longitudinal experimental study focusing on Twitter users with positive discussions on immigration and anti-immigrant sentiments revealed that the most persuasive effect was achieved when a social robot initially posted content opposing immigration and gradually shifted to more supportive content. This strategy of initially agreeing with the audience, following the audience's emotions, establishing a connection, and then stating its own target viewpoint proved more effective than other strategies, suggesting that presenting nuanced and gentle arguments based on the original language of the audience helps people encounter opposing views while mitigating the "backfire effect.".[11]

2.2. Collective Behavior Level

The social robots themselves possess a social nature. They engage in self-organization, relying on their own behavioral logic and observing the environment in which they are situated to engage in local interactions with nearby counterparts.[12] Consequently, they emerge in the form of groups or networks of robots.[13] Typically, a "machine cluster network" will emerge complex properties and higher-order functionalities that individual machines do not possess.[14] In a collective manner, these networks further engage in discourse interaction practices with the public, political organizations, media, and other social robots. Through organized dissemination, they contend for discursive authority within the public opinion ecosystem.[15]

Machine collectives primarily engage in the construction of online public opinion through two main methods. The first approach involves the collective forwarding of content to collectively create machine opinion leaders, thereby expanding the influence of central node accounts in the machine network. For example, analyzing the tweet forwarding user structure of two typical social robots, "@Indddy77" and "@Aditya_Sindh," driving hashtag activities during the Russia-Ukraine conflict, it was found that out of 1003 users involved in forwarding, 106 were social robot accounts. These two robots formed a strong network connection through a set of commonly forwarding accounts.[16] The second approach involves collectively promoting specific content to hinder the circulation of diverse viewpoints, creating a "spiral of silence" effect to enhance the influence of the represented positions and viewpoints. Combining with the collaborative filtering mechanism of recommendation systems, social robots can induce the spread of specific information. Moreover, a large number of social robots continuously propagate the same viewpoint throughout the day. Due to a lack of checks and balances, they unintentionally contribute to constructing a "conformity environment" in public opinion. This reduces the chances for netizens to encounter views from other perspectives, potentially creating a deceptive appearance for one side and influencing public judgment through the spiral of silence effect, posing a serious threat to democracy.[17] Research indicates that social robots only need to account for 5%-10% of discussants to alter public opinion and make the viewpoints they propagate eventually dominate. [18]

DOI: 10.54254/2753-7048/50/20240802

2.3. Hybrid Human-Machine Behavior Level

Intelligent media technology integrates heterogeneous communication entities with different interest appeals into a seamlessly intertwined temporal and spatial context.[19] Machines can connect with and shape other "species" in the discourse ecosystem, forming a hybrid human-machine network.

The first pattern of social robot participation in the construction of online discourse at the hybrid human-machine behavior level primarily reflects its influence on human behavior in the discourse ecosystem. A study exploring the interactive mechanisms among social robots, the public, and the media in social networks concerning the China-U.S. trade war reveals that in the first-level agenda-setting, the agenda of social robots positively impacts the public agenda. In the second-level agenda-setting, mutual influence exists, and within the substantial attributes of the second level, media is influenced by social robots in negative emotional attributes.[20] The second pattern of social robot participation at the hybrid human-machine behavior level manifests in the coordinated, cooperative, and competitive behaviors autonomously formed by humans and machines to drive discourse in a specific direction. Regarding the coordination and cooperation between humans and machines, social robots and the public collaboratively promote hashtag activities to create trending topics. Additionally, social robots actively engage in positive interactions with humans, initiating conversations through "@other accounts" and even amplifying the influence of original tweets by retweeting content from human opinion leaders. Furthermore, social robots and media may mutually benefit each other. When the positions of social robots align with those of the media, they may actively retweet media content or links to amplify its reach. Research suggests that media bots act as "early spreaders." The moment an article link is posted on a highly credible media website, media bots begin their activities, potentially contributing to the viral spread of media content and becoming an opportunity for social media trends.[21]

Despite comprehensive discussions in existing research on social robots, a shared limitation is the lack of in-depth analysis on algorithmic gender bias as a widespread societal phenomenon. This study, grounded in the paradigm of machine behavior studies, focuses on investigating gender bias issues inherent in algorithms.

3. Presentation of Algorithmic Gender Bias under the Paradigm of Machine Behavior Studies

In the era of artificial intelligence, algorithms are intricately linked with machine behavior. Algorithms dictate rules for machine data processing, instructing machines to process data in specific steps. Machines continuously optimize models through learning and analyzing the connections between data, eventually forming mature models and identifying optimal solutions. [22] However, these algorithmic models are not flawless. During the modeling process, issues such as data bias, learning frameworks, classification criteria, and evaluation metrics may introduce defects. These defects result in errors and deviations in terms of reliability, accuracy, and precision, leading to discrepancies between computational results and real-world situations. In some cases, the outcomes may be unfavorable for certain parties, revealing a certain degree of "bias." To explore the characteristics of algorithmic gender bias under the paradigm of machine behavior, this paper will follow the three perspectives of "individual machine behavior - collective machine behavior - hybrid human-machine behavior." The analysis framework will unfold the research.

3.1. Individual Machine Behavior

Research on individual machine behavior typically focuses on specific intelligent machines. It usually emphasizes the intrinsic attributes of individual machines, often driven by their source code or design.[1] However, the emergence of new technologies inevitably carries the genes of human society,

DOI: 10.54254/2753-7048/50/20240802

embedding itself in all elements that construct what we call the human social framework. [23] Within these elements, structural biases of human society may be embedded in algorithmic practices. Therefore, in individual machine behavior, the expression of algorithmic gender bias manifests as: interference of human bias in the collection of raw data and the establishment of databases, and the permeation of human bias in algorithm program design.

3.1.1. Selective Presentation of Gender in Raw Databases

Fundamentally, data is a technical medium for the reproduction of human representation, constrained by the reality of the medium and objective reality. The expressed data may not necessarily reflect the truth, and the unexpressed data may not necessarily be nonexistent. Therefore, assembling reality from data inevitably involves data biases. If gender-related data represents only a portion of the population, bias is difficult to avoid. Taking online shopping as an example, women are the primary participants, but deducing the traditional notion of "men produce, women consume" from female shopping data introduces bias. This bias arises because data on women participating in production is not included in the statistics. From the perspective of data collection purposes, data exhibits a strong e-commerce bias, where each user login and click is designed to represent a potential purchase behavior, motivating a tendency to generalize based on parts.[4]

Secondly, structural biases in raw databases lead to the replication of biases in intelligent algorithms. The original data for algorithms comes from human society and inevitably carries the ideology of that society. Researchers identified a bias in AI programs when recognizing a set of photos: individuals depicted wearing aprons in the kitchen were consistently identified as female, regardless of actual gender. The reason for AI errors lies in the training data used for image recognition, where ninety percent of the images associate women with the kitchen. As AI learns and trains on such biased databases, it naturally reproduces these biases in image recognition.[24] Thus, to extract and analyze data and information from the real world, algorithmic technology inevitably replicates structural biases embedded in the original databases of the real world, subsequently influencing the operation and results of algorithms.

3.1.2. Bias Loop in Program Design

During the development process, algorithms are susceptible to the subjective influences of the designers. If designers harbor gender bias, value discrepancies, or lack gender sensitivity, their subjective inclinations in coding rules and metric selection can lead to the generation of algorithmic bias. Taking the example of the domestic online knowledge community and paid course platform "QianLiao," it offers paid knowledge categories covering a wide range, including maternity and infant care, parenting, health, medicine, wellness, beauty, emotions, family, workplace, and human resources, among others. There is evident gender differentiation in content marketing. Content targeting males emphasizes practicality and has higher entry barriers, while content targeting females prioritizes subjective experiences, focusing on health, wellness, beauty care, and emotional aspects. Page data indicates that among the top-ranked courses for males are those related to workplace, finance, automobiles, history, and technology – "hard" knowledge topics. On the other hand, among the top 10 courses for females are topics such as postpartum shaping, yoga, skincare, beauty, etiquette, and workplace fashion – "soft" knowledge topics. In terms of algorithmic models and logic, whether utilizing collaborative filtering algorithms based on user identity or information elements, or employing regression analysis, the user labels or user profile strategies defined by these algorithms tend to carry gender stereotypes. This constitutes a form of algorithmic bias imbued with knowledge power hues.[25]

DOI: 10.54254/2753-7048/50/20240802

3.2. Collective Machine Behavior

In comparison to individual machine behavior, research on collective machine behavior regarding algorithmic gender bias predominantly focuses on the interaction within machine clusters and the behavioral patterns at the system level.[1] The primary analysis involves how these machines mutually influence each other, form specific machine networks, and the interaction and propagation patterns between different machine networks. In some cases, individual machine behavior may not offer valuable insights for analyzing clustered machine behavior.

3.2.1. Creating Opinion Leaders to Generate Information Fog

Once robots interacting locally through simple algorithms aggregate into a large-scale collective, intriguing behaviors emerge.[1] A batch of robots can create the illusion of tweet popularity through joint retweeting, forming opinion leaders that influence others' perceptions of the account. The presence of opinion leader-type social bots was identified in discussions related to the Russia-Ukraine conflict, exemplified by the account @UAWeapons. Analyzing all tweets from this account and examining the retweet network of the 15 most retweeted tweets revealed that among the 1,045 accounts with four or more retweets, 447 were identified as social bots, accounting for 42.78%. Thus, social bots constitute a significant proportion in the retweet network of @UAWeapons, serving as a crucial variable propelling the account's rapid popularity.[16] Due to algorithms neglecting pseudo data in information hotness, biased evaluations of information recommendation value can occur. During the same monitoring period, information with high click-through rates and retweet volumes is generally prioritized by the algorithm as the top topic trend or pushed to users first. Consequently, algorithms are deceived by false traffic, setting agendas for users in the information recommendation mechanism, reinforcing the popularity of "hot information," while "less popular information" goes unnoticed, perpetuating a cycle of information bias. [26] Machine cluster behavior dominated by opinion leaders exhibits characteristics of scalability, purposefulness, and misguidance, serving as the primary entity in creating information fog. The "objectivity" they display enhances the permeation, deception, concealment, difficulty of accountability, and deeper harm of the concealed "gender discrimination against women."

3.2.2. Large-Scale "Robot Water Army" Playing a Negative Role

Search engines employ algorithms to develop the "Auto-Complete" feature, where, upon entering keywords in the search box, users are automatically shown or completed with relevant text based on the historical search volume of the terms. The initial purpose is to reduce users' time costs for online searches. However, biases implicit in keyword text from historical searches can significantly impact algorithmic objectivity and user perception.[26] Former German First Lady Bettina Wulff sued Google because the search engine's "Auto-Complete" results associated her name with terms like "prostitute" and "escort girl." [27] Google was ordered to modify the "Auto-Complete" results of its search engine. The presentation of such severely biased results by the search engine is closely related to the search volume generated by the army of robots recruited by Craigslist and Amazon Turkey. These highly organized and large-scale "robot water armies" play a negative and misleading role closely tied to the production, dissemination, and consumption of fake news. [28] There are currently companies both domestically and internationally specializing in creating and selling "robot water armies" on social networks. Through these "robot water armies," the spread of false information can influence advertising audiences, damage corporate reputations, and sway public decisions.

Nowadays, many organisms, as well as an increasing number of AI intelligent machines, are complex entities with behaviors or interactions that may defy simple characterization. When entities

DOI: 10.54254/2753-7048/50/20240802

engaged in interactions can exhibit complex cognition, the emergence of additional properties remains a key challenge in the field of biological sciences. [1]

3.3. Hybrid Human-Machine Behavior

In today's world, the relationship between humans and machines is becoming increasingly intimate. In the interaction between the two, machines not only shape the information we receive but also establish relationships with us sufficient to change our societal systems. Due to their complexity, these hybrid human-machine systems constitute one of the most challenging yet crucial research areas in machine behavior. [1]

3.3.1. Artificial Intelligence Shapes Female Images, Constructs Female Cognitive Thinking

The shaping of human behavior by machines refers to the way intelligent machines introduced into social systems can alter human beliefs and behaviors.[1] The image projected onto robots is a reflection of human desires. In the subtle infiltration of artificial intelligence products, the female images shaped by artificial intelligence quietly "construct female cognitive thinking." [29] CEO of Live Person, Robert LoCascio, once mentioned, "I accidentally overheard my two-year-old daughter talking to Amazon's voice assistant, Alexa... Alexa set a bad example for my daughter — that women should be submissive, tolerate disrespectful behavior, and stay at home." [30] Moreover, concerns arise about how artificial intelligence products also "shape" women. On one hand, a variety of photo editing tools and beauty cameras have become essential items for modern women. A 2016 report from the Aurora Research Institute indicated that nearly 90% of users for daily photo editing apps are female, and 83.1% of beauty camera users are women, predominantly young users. [31] The trendy "snake face" and "mixed-race face" on social media have become standards for retouching, causing the loss of individuality. On the other hand, following the popular images of women on social media, cosmetic surgery has become a trend, lowering the age limit. This not only distorts women's aesthetic values, harms their bodies but also leads to self-loss, leaving only the homogeneous female image mass-produced by "mechanized tools."

3.3.2. Algorithms Learn Human Biases in Interaction with Users

Intelligent machines can change human behavior, and humans can create, influence, and shape the behavior of intelligent machines. [1] Therefore, information in social interactions and user emotional tendencies will also affect algorithm training. Chatbots can learn gender biases and discrimination from social interactions. Artificial intelligence products, while catering to user needs, filter out information that users do not need, reinforcing existing user tendencies, even if these tendencies are problematic. [32] Humans shape machine behavior through direct operation of AI systems and through active training and passive observation of data generated by human behavior. [1] Studies on Artificial Moral Agents (AMAs) found that the process of machines learning human language skills deeply absorbs various biases implicitly present. Complete artificial moral agents are yet to emerge, and existing machines cannot consciously resist biases. [33] Microsoft's Twitter chatbot Tay was urgently taken offline less than a day after interacting with users because, during the conversation with humans, Tay was taught to be an extreme individual who spewed profanity, exhibited gender and racial discrimination. Existing artificial intelligence machines do not possess the ability to automatically recognize and resist human biases. Therefore, in the process of human-machine interaction, machines unconsciously and indiscriminately learn all ethical aspects and preferences of humans.

DOI: 10.54254/2753-7048/50/20240802

3.3.3. Social Robots Amplify Gender Bias Issues Through Hybrid Human-Machine Behavior

Human-machine collaborative behavior refers to the role played by most artificial intelligence systems in complex hybrid systems coexisting with humans. Research on these systems manifests as behavioral features of human-machine interaction, such as cooperation, competition, and coordination—all of which are issues of significant research significance. [1] Regarding the coordination and cooperation between humans and machines, social robots and the public collaboratively promote hashtag activities to create trending topics. For example, in the mixed public opinion ecology of the Russia-Ukraine conflict on Twitter, there is an abundance of sharply positioned “hashtags” supporting either Russia or Ukraine. Originally initiated by humans, these “hashtags” can be amplified by social robots through retweets, repeated pushes of the same content within tags to increase attention, and other methods to quickly propel the “hashtags” to the top trending list, garnering broader public attention and recognition resources. [16] Additionally, there is mutual utilization between social robots and users. Creators of fake news rely, to some extent, on an online robot network that supports fake news websites. [34] Social robots not only act as producers of fake news but also serve as disseminators and consumers, weaving a network for the spread of fake news together with real users. Social robots are among the most active accounts spreading fake news in internet communities. They “can change the dynamic structure of social networks through an artificial intelligence-driven model to enhance the efficiency of dissemination.” [35] Humans manipulating social robots make these robot clusters multiple sources of fake news infection in social networks. Other social robots and real social media users, by obtaining and sharing fake news, become infected, transforming into lower-level sources of infection participating in the amplification path of fake news dissemination. Currently, in the field of human-machine collaborative behavior, there is a situation where social robots on social media use topics that create gender conflicts, leading to disputes between the sexes. If manipulated or exploited by entities or organizations holding extreme attitudes, this can magnify societal gender biases, provoke gender conflicts, and increase the stereotypical impact of audiences on gender discrimination.

4. Conclusion

In the face of gender discrimination issues brought about by algorithms, technical aspects such as algorithm transparency, algorithm free competition, and algorithm substitution cannot solve the problem. This is because “complex technology itself is only a means; it should obey and serve a higher and distant purpose.” [36] Algorithmic gender discrimination is the result of the interaction and construction of society and technology. It requires us to start from the perspective of social construction, considering the relationship between society, gender, and algorithms. A more proactive policy intervention is needed to prevent the “coding gaze” in algorithm use and decision-making.

The rapid development of algorithmic technology and artificial intelligence poses a threat to the consensus vision of gender equality, emphasizing the urgent need for governance of algorithmic gender bias. Brian Arthur, in “The Nature of Technology,” discusses that, whether we have noticed it or not, technology has become oppressive and troublesome at this historical stage. [37] The relationship between humans and technology is no longer a one-sided power suppression, but rather a sustainable development of a community where humans and technology mutually constrain and promote each other. The ultimate goal of governing algorithmic gender bias is to combat gender bias in the digital and technological realms, ensuring that different genders maintain their inherent value and dignity in human-machine interactions. It aims to extend gender equality in the era of intelligent communication rather than suppress it.

References

[1] Rahwan, I., Cebrian, M., Obradovich, N., et al. (2019). Machine behaviour. Nature, 568(7753), 477-486.
[2] Hyunjin Kang, & Chen Lou. (2022). AI Agency vs. Human Agency: Understanding Human-AI Interactions on TikTok and Their Implications for User Engagement. Journal of Computer-Mediated Communication, 27(5), 1.
[3] West, M., Kraut, R., & Ei Chew, H. (2019). EQUALS y Unesco—2019—I'd Blush If I Could: Closing Gender Divides in Digital Skills through Education. UNESCO: Paris, France. 144p.
[4] Meyer, D. (2018, October 10). Amazon Reportedly Killed an AI Recruitment System Because It Couldn't Stop the Tool from Discriminating Against Women. Fortune. http://fortune.com/2018/10/10/amazon-ai-recruitment-bias-women-sexist/.
[5] Zhang H., Duan Z., & Han X.. (2019). "Differentiation or Coexistence: Exploring the Research Path of Social Robots in Social Media." Journalism, 2019(2), 10-17.
[6] Wang P., Huang W., & Cao B.. (2020). "Integration and Differentiation: A Temporal Analysis of the Evolution of Diverse Subject Opinions on Weibo under the Epidemic." Journalism University, 2020(10), 16-33, 118-119.
[7] Zhang T. (2006). "Expression of Public Opinion in the Context of Media Society." Modern Communication (Journal of Communication University of China), 2006(5), 12-15.
[8] Zeng F. & Wang Y.. (2015). "The Influence of Social Media on Celebrity Public Action." Exploration and Contention, 2015(12), 72-76.
[9] Zhang H., Duan Z., & Yang H.. (2019). "Analysis of Political Robot Opinion Intervention in the Social Media Space." Journalism, 2019(9), 17-25.
[10] Heidari, M., James Jr, H., & Uzuner, O. (2021). An empirical study of machine learning algorithms for social media bot detection. 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 1-5.
[11] Yang, Q., Qureshi, K., & Zaman, T. (2021). Mitigating the backfire effect using pacing and leading. International Conference on Complex Networks and Their Applications, Springer, Cham, 156-165.
[12] Zheng, C., & Fan, H. (2020). "From Social Contagion to Social Diffusion: Research on the Social Diffusion Mechanism of Social Robots." Journalism, 2020(3), 51-62.
[13] Assenmacher, D., Clever, L., Frischlich, L., et al. (2020). Demystifying social bots: On the intelligence of automated social media actors. Social Media + Society, 6(3), 2056305120939264.
[14] Rahwan, I., Cebrian, M., Obradovich, N., et al. (2019). Machine behaviour. Nature, 568(7753), 477-486.
[15] Chen, L. (2018). "Control and Prevention of Opinion Entropy: A Methodological Understanding of Network Governance." Journalism and Communication Research, 2018(8), 65-80+127.
[16] Zhao, B., Zhang, H., Ren, W., Zhang, Y., & Liu, S. (2022). "Tags, Accounts, and Narratives: Research on Opinion Intervention of Social Robots in the Russia-Ukraine Conflict." Journalism and Writing, 2022(9), 89-99.
[17] García-Orosa, B. (2021). Disinformation, social media, bots, and astroturfing: The fourth wave of digital democracy. Profesional de la información, 30(6).
[18] Cheng, C., Luo, Y., & Yu, C. (2020). "Dynamic Mechanism of Social Bots Interfering with Public Opinion in Network." Physica A: Statistical Mechanics and its Applications, 551, 124163.
[19] Gao, X. (2019). "Research on the Influence of Intelligent Media Technology on the Evolution of Mainstream Opinion." Modern Communication (Journal of Communication University of China), 2019(5), 5-11.
[20] Zhao, B., & Zhang, H. (2022). "Issue Shifting and Attribute Highlighting: Research on the Agenda Setting of Social Robots, the Public, and the Media." Journal of Communication and Sociology, 2022(59), 81–118.
[21] Santini, R. M., Salles, D., Tucci, G., et al. (2020). Making up audience: Media bots and the falsification of the public sphere. Communication Studies, 71(3), 466-487.
[22] Liu, Y. H. (2019). "Algorithmic Bias and Its Regulatory Path: A Study." Law Journal, 2019(6), 55-66.
[23] DeNardis, L. (2017). The Global War for Internet Governance (Q. Qingling & C. Huihui, Trans.). China Renmin University Press, p. 10.
[24] When AI Learns Gender Discrimination. Huxiu Web. [Online]. Available: https://www.huxiu.com/article/258405.html?h_s=f1.
[25] Fan, H. X., & Sun, J. B. (2021). "The Invisible 'Elephant': Gender Discrimination in Algorithms." News Enthusiasts, 2021(10), 29-32. DOI:10.16017/j.cnki.xwahz.2021.10.08
[26] Guo, X. P., & Qin, Y. X. (2019). "Deconstructing the Data Myth of Intelligent Communication: Causes and Risk Governance Paths of Algorithmic Bias." Modern Communication (Journal of Communication University of China), 41(09), 19-24.
[27] Morozov, E. (2014). To Save Everything, Click Here: The Folly of Technological Solutionism (X. Zhang & J. Lv, Trans.). Electronic Industry Press, pp. 156, 167, 152.
[28] Hu, H. C., & Xie, X. Z.. "Development Trends and Governance Issues of Fake News under the Background of Artificial Intelligence." News Enthusiasts.

[29] Wang, H. J. (2020). "Gender Discrimination in the Consumption Scene of Artificial Intelligence." Dialectical Naturalism Newsletter, 42(05), 45-51. DOI:10.15994/j.1000-0763.2020.05.007.

[30] LoCascio, R. (2018, May 11). 'Thousands of Sexist AI Bots Could be Coming. Here's How We Can Stop Them.' [Online]. Available:

[31] Aurora Big Data Research Institute. (2016, December 12). Research Report on Photo Editing Apps [Research Report]. http://www.199it.com/archives/545479.html

[32] Yan, J. (2019). "Algorithmic Discrimination and Countermeasures in Artificial Intelligence." Rule of Law Review, 2019(4).

[33] Wallach, W., & Allen, C. (2017). Moral Machines: How to Build Ethical Robots (X. Wang, Trans.). Peking University Press.

[34] Tandoc Jr, E. C. (2019). "The Facts of Fake News: A Research Review." Sociology Compass, 13(9), e12724.

[35] Zheng, C. Y., & Fan, H. (2020). "From Social Contagion to Social Diffusion: A Study on the Social Diffusion Mechanism of Social Bots." Journalism, 2020(3), 51-62.

[36] Manjoo, F. (2017). True Enough: Learning to Live in a Post-Fact Society (J. Song, Trans.). Triple-A Publishing, p. 127.

[37] Arthur, B. (2018). The Nature of Technology (D. Cao, & J. Wang, Trans.). Zhejiang People's Publishing House, pp. 5, 36, 11, 41.