Technical Governance Measures Against Misinformation on Foreign Platform Media and Their Implications (Postprint)

Authors: Shen Jinxia, Zhang Jiaying

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

Abstract

Currently, cyberspace governance has emerged as a critical global agenda. Disinformation propagates virally on platform-based media, and there is a growing consensus on strengthening platform accountability and developing technological governance mechanisms. To harness technological capabilities for detecting massive volumes of automatically generated and disseminated disinformation, foreign platform-based media have actively adopted advanced technologies in recent years to combat platform disinformation, including: labeling to enhance disinformation identifiability; employing AI for automated detection and annotation of disinformation; utilizing algorithms to elevate authoritative information rankings or suppress disinformation rankings; and leveraging machine learning to detect images, videos, and deepfakes, as well as to identify social bots. This article systematically reviews relevant foreign technological governance measures to provide reference and guidance for China' s platform-based media in combating disinformation through technological means.

Full Text

Technical Governance Measures Against Misinformation on Foreign Platform-Based Media and Their Implications

Shen Jinxia, Zhang Jiaying
(Communication University of China, Beijing 100024)

Abstract: Internet governance has become a critical global issue. As misinformation spreads virally on platform-based media, strengthening platform accountability and developing technical governance mechanisms has become a consensus. To leverage technological power in detecting massive volumes of automatically generated and disseminated misinformation, foreign platform-based

media have actively adopted advanced technologies in recent years. These include labeling to increase misinformation identifiability, AI-driven automated detection and annotation, algorithmic elevation of authoritative information rankings or demotion of misinformation, machine learning detection of images, videos, and deepfakes, and identification of social bots. This article systematically reviews current foreign technical governance measures to provide reference for Chinese platform-based media in combating misinformation through technological means.

Keywords: misinformation; platform-based media; deepfake; social bots; platform governance

Classification: G210
Document Code: A
Article ID: 1671-0134(2021)08-007-06
DOI: 10.19483/j.cnki.11-4653/n.2021.08.001

Introduction

The concept of "Platisher" (platform-based media) was first proposed by Jonathan Glick in his article *Rise of the Platishers*, combining the strengths of platforms and publishers. A Digiday contributor further defined it as "entities that possess both the professional editorial authority of media and an open, user-facing digital content platform" [1]. Websites such as BuzzFeed, Medium, and Toutiao, as well as social media platforms including Facebook, Twitter, YouTube, TikTok, and WeChat, all exhibit characteristics of platform-based media. These platforms integrate specific algorithmic technologies with professional editorial operations, generating enormous influence in content production, aggregation, and distribution. Consequently, platform-based media embody both the open nature of technology platforms and the gatekeeping attributes of media publishing [2].

Platform-based media have become the primary channel through which users access information. However, they have also become breeding grounds for clickbait, misinformation, and vulgar content, which persistently plague online discourse and internet governance. The spread of misinformation on these platforms has grown increasingly complex. Visual and audio-visual misinformation spreads rapidly, and its copies are difficult to completely eliminate. During the 2020 U.S. election, a video originally posted on TikTok allegedly showed a California voter filling out seven ballots for Democratic candidate Joe Biden. Although the creator clarified these were sample ballots and the video was removed for violating TikTok' s community guidelines, copies continued to circulate widely on Twitter and Facebook as evidence of electoral fraud [3]. Social bots, as automated accounts, can amplify information dissemination by orders of magnitude. In recent years, Twitter bots have been exploited by political groups to spread misinformation and influence public opinion in political events such as U.S. elec-

tions [4]. Deepfake videos, such as those showing Obama insulting Trump or Mark Zuckerberg boasting about "completely controlling stolen data of billions of people,"have attracted widespread attention. Deepfakes achieve self-evolution through adversarial training of two neural networks, making generated content difficult to authenticate.

The reality has forced platforms to assume greater responsibility, making platform governance a critical issue in global journalism and communication research. Current Chinese research on misinformation in platform-based media focuses either on analyzing dissemination mechanisms in specific cases or discussing government regulation, policy formulation, and user media literacy, while the technical dimension of platform governance requires deeper exploration. Therefore, systematically reviewing foreign technical governance measures holds significant reference value for China' s internet content governance.

---

1. Technical Governance Measures Against Misinformation on Foreign Platform-Based Media

To combat misinformation and enhance overall platform credibility, major foreign platform-based media have implemented technical governance measures including labeling, ranking adjustments, and AI detection.

1.1 Labeling + AI-Powered Automated Detection Labeling refers to the practice of attaching additional information to user-generated content, including fact-checking results, content warnings, or contextual information, enabling users to assess information authenticity based on these indicators. The most common types are "credibility labels" and "contextual labels" [5]. Credibility labels provide explicit information about information veracity. Since March 2020, Facebook has used such labels to mark over 180 million posts containing election-related misinformation [6]. The platform covers original posts with a gray background and adds a "False Information" warning label, directly informing users that the post has been fact-checked as false and providing a link to detailed debunking information. Contextual labels merely provide additional background without making explicit credibility claims, encouraging users to make informed judgments after considering more context. In December 2020, TikTok launched the #covidvaccine hashtag to detect and label all COVID-19 vaccine-related videos, attaching a banner reading "Learn more about COVID-19 vaccines" that directs users to verifiable, authoritative sources [7]. Different platforms vary in their terminology, placement, visual design, and interaction design for labels. Twitter categorizes misinformation labels as misleading information, disputed claims, and unverified claims, applying different measures such as removal or warnings accordingly.

Platform-based media combine third-party fact-checking with automated technology to detect and label massive information volumes. Facebook launched

its fact-checking program in December 2016, hiring third-party organizations to evaluate content accuracy. By October 2020, Facebook had partnered with 80 fact-checking organizations covering over 60 languages globally [8]. User reports or expressions of doubt about Facebook posts serve as signals for third-party fact-checkers to intervene. Content identified as misinformation by fact-checkers receives a "false information" label. Facebook uses previously fact-checked articles as training data, employing machine learning to expand detection scope and overcome the limitations of manual review in terms of volume and speed. In April 2020, based on approximately 7,500 articles reviewed by fact-checking organizations, Facebook used AI to label about 50 million COVID-19-related pieces of content [9]. Twitter' s team also claims to be using and improving systems to rapidly detect and label COVID-19-related content while ensuring labels themselves do not amplify misinformation dissemination. Instagram uses image-matching technology to find similar content and add labels. Additionally, Facebook employs SimSearchNet, a specialized near-duplicate detection algorithm, to identify copies of image-based misinformation such as screenshots. Through precise matching, SimSearchNet enables identification and labeling of misinformation copies [10].

Research demonstrates that labeling effectively curbs misinformation spread. A 2019 University of California study found that labeling false information influences users' sharing intentions, thereby reducing dissemination [11]. Contextual labels providing background explanations and authoritative content help users reconcile cognitive dissonance with facts and promote more lasting belief changes [12]. However, effective misinformation reduction requires large-scale coverage.

Platform-based media like Facebook and Twitter are also using AI and other technologies to train systems for automatic detection and labeling of massive information volumes, overcoming manual verification limitations.

1.2 Prioritization and Demotion For platform-based media, algorithmic architecture determines content aggregation and distribution, representing a key distinction from traditional news websites' linear presentation. To achieve personalized customization and avoid information overload, platforms employ recommendation algorithms to filter information, evaluating, ranking, and pushing content according to platform-defined rules. This algorithmic decision-making process comprises four steps: prioritization, classification, association, and filtering .

Different platforms assign varying weights to algorithmic elements. Facebook' s News Feed, its core news distribution project, has evolved from the EdgeRank algorithm to incorporating user relationships, preferences, and recent contacts, then to collaborative filtering mechanisms, demonstrating that Facebook' s content recommendation is based on users' social relationships and interaction preferences. In contrast, Google' s recommendation logic prioritizes content categorization and quality, while TikTok' s mechanism relies more on individual user preferences and activity history, including likes, shares, and follows [13].

Despite these differences, all platforms share a common principle: more clicks lead to more recommendations.

Algorithmic recommendation systems on YouTube, Facebook, and Twitter bear significant responsibility for spreading misinformation, hate speech, conspiracy theories, and other harmful content. Misinformation often associates with high-attention social events or political issues and carries emotional sensationalism, making it more likely to attract user engagement. Once entering recommendation systems, such information gets recommended to more users through social connections, proximity, popularity, and related topics, creating viral spread through automated algorithmic decisions.

The solution involves reducing misinformation' s entry opportunities into recommendation systems through downranking or demotion, decreasing its visibility. After WHO declared COVID-19 a global health emergency, platforms actively 采取措施减少虚假信息和有害内容的传播. In August 2020, Facebook publicly disclosed its content recommendation operations, stating that clickbait, deceptive information, and false or misleading content would be demoted, and users/groups frequently sharing misinformation would receive reduced recommendations [15]. To combat COVID-19 infodemic, Facebook demotes misinformation identified by third-party fact-checkers and WHO, including "exaggerated or sensational health claims and those using health claims as pretexts for product/service sales," while elevating authoritative information rankings. Google similarly reduces rankings for content identified as containing misinformation while prioritizing official sources. Instagram removes such content from "Explore" and hashtag pages and reduces its visibility in feeds and stories. Through demotion, misinformation' s ranking in information streams is lowered or eliminated, significantly reducing its chances of entering recommendation systems and suppressing its reach.

1.3 Machine Learning Detection of Images, Videos, and Deepfakes In the visual communication era, users consume massive amounts of visual content daily. YouTube, as a user-generated video community, ranks second globally after Facebook in user activity [16]; TikTok has 689 million monthly active users worldwide and became the second most-downloaded iPhone app in 2020 [17]. As user behavior evolves, misinformation has developed from "unimodal" to "multimodal" forms. According to Hazel Baker, Reuters' global head of user-generated content, "In almost every major global news event in 2019, misleading videos and images were identified on social media" [18].

In this context, human-driven fact-checking faces significant limitations, as manual verification struggles to discern authenticity and lags far behind misinformation' s propagation speed and volume.

Facebook expanded fact-checking to images and videos in 2019, categorizing visual misinformation into three types: manipulated or fabricated, out of context, and text or audio claims. Out-of-context content uses images or videos

divorced from their original context, distorting their authenticity. Fact-checkers can identify this type using journalistic expertise and contextual understanding. The first and third types are more complex, requiring technical solutions for large-scale detection.

For photos containing false text or audio, the first step involves using Optical Character Recognition (OCR) or audio transcription tools to extract text, then applying natural language processing to match extracted text against fact-checker-verified misinformation databases for duplicates. Facebook built Rosetta, a large-scale machine learning system that extracts text from over one billion public images and video frames daily on Facebook and Instagram, feeding it into trained text recognition models to help machines understand text-image composition and context [19]. This technology enables Facebook to conduct variable and large-scale image analysis.

Manipulated images and videos are most difficult to identify, with deepfakes being one such type. Deepfakes use Generative Adversarial Networks (GANs), where "generators" and "discriminators" compete in a learning process through continuous parameter adjustment. The ultimate goal is for discriminators to be unable to judge generator outputs' authenticity, making deepfake images and videos often indistinguishable from reality. To better detect deepfakes, Facebook partnered with Michigan State University' s research team to identify and track deepfake information through reverse engineering. The system first assumes an image is a deepfake, then uses reverse engineering algorithms to trace the AI software that created it. A fingerprint estimation network runs on deepfake images to evaluate fingerprints left by AI software. Once the fingerprint is known, it can be matched against deepfake video/image fingerprints in algorithmic outputs. This approach can trace which specific software generated the deepfake product. The system achieves 70% accuracy in key benchmark tests, outperforming previous detection methods [20].

1.4 Machine Learning Detection of Social Bots Social bots are algorithmic programs that automatically generate content, mimicking and influencing human behavior while interacting with users on social networks. Initially used for content aggregation and automatic replies, they have been maliciously designed to spread misinformation, send spam, or simply create noise to mislead and manipulate public opinion [21]. Research indicates social bots constitute 9% to 15% of Twitter' s active users [22]; Instagram may have 95 million bot accounts, representing 9.5% of total users [23]. Compared to humans, social bots are more active in sharing—one study analyzing 1.2 million tweets with hyperlinks found approximately 66% of links from popular news media sites were automatically posted by bot accounts [24].

The danger of social bots lies in their unchecked content dissemination, including various forms of misinformation and hate speech. Thousands of coordinated bots drive exponential misinformation growth on platforms. In recent years, social bots' information amplification function has been maliciously exploited,

threatening the online ecosystem. They are believed to have influenced online discussions during major Western political elections, including the 2016 U.S. presidential election and Brexit referendum [25]. One study found that since January 2020, bot accounts accounted for up to 45% of Twitter accounts discussing COVID-19 information, fabricating over 100 false narratives about the virus, including conspiracy theories about hospitals using mannequins as patients to exaggerate severity and claims linking COVID-19 transmission to 5G towers [26]. Therefore, identifying and removing these malicious social bots is crucial for platforms.

Social bot detection technologies primarily include graph-based, feature-based, and crowdsourcing methods. Crowdsourcing relies on professional manual identification, while the first two employ machine learning [27]. Graph-based detection analyzes platform user relationships through account social network graphs. Real users exhibit substantial following, retweeting, and bidirectional interaction behaviors, resulting in network graph structures significantly different from bot accounts [28]. Feature-based detection analyzes account behavior patterns from multiple perspectives. The Atlantic Council' s Digital Forensic Research Lab (DFRLab) proposes identifying political social bots through 12 signals including account activity, anonymity, and posting behavior characteristics—such as minimal original content, high retweet volumes, absence of comments on retweets, and tweets containing multiple languages [29]. Botometer, developed by researchers at the University of Southern California and Indiana University, is a typical feature-based tool that reads over 1,000 account features and assigns a score between 0 and 1, with higher scores indicating greater likelihood of being a social bot.

Facing massive numbers of social bots and misinformation, Twitter continuously improves its automated detection technology. In 2019, Twitter acquired Fabula AI, a company specializing in machine learning and misinformation detection. The company excels in geometric deep learning, which applies machine learning to network-structured data to describe large, complex relational and interaction datasets [30], helping Twitter better improve its information environment. Additionally, Twitter requires users to complete simple reCAPTCHA processes or password reset requests– "Completely Automated Public Turing tests to tell Computers and Humans Apart" –that require users to operate on text images unreadable by OCR software, thereby identifying social media bots. Research teams studying social bots' impact on U.S. elections in 2016 and 2020 noted that bot numbers have decreased dramatically compared to 2016, likely due to platforms like Twitter achieving success in detecting and removing social bots.

---

2. Implications for Chinese Platform-Based Media Governance

The COVID-19 pandemic and concurrent infodemic in 2020, compounded by U.S. election-related misinformation, have made cyberspace governance increas-

ingly urgent. Global consensus has emerged on strengthening platform accountability, developing technical governance mechanisms, and improving laws and regulations to promote collaborative governance against misinformation and extremist content. For China, internet governance has long been a significant topic. Internet application and popularization have created new information dissemination environments, facing internal challenges from negative public opinion rooted in social contradictions and identity conflicts, and external challenges from international cyberspace discourse power competition, cybersecurity, and online ideological struggles [32]. In this context, China has continuously explored internet governance pathways. President Xi Jinping' s report at the 19th Party Congress emphasized "strengthening internet content construction, establishing comprehensive internet governance systems, and creating a clean cyberspace," reaffirming the importance of internet governance. With the arrival of 5G and the intelligent era, platform-based media will possess more powerful computing capabilities and resources, making them better positioned and obligated to function as "online gatekeepers" [33]. Foreign platform governance measures offer valuable insights and references.

First, continuously optimize and adjust algorithms by "prioritizing" or "demoting" different content. Misinformation, clickbait, pornography, and vulgar content are focal criticisms of platform-based media, prompting algorithmic adjustments. Scholars note that Toutiao' s recommendation algorithm has undergone four major adjustments and upgrades since its first version, manifesting in semi-automatic human-algorithm collaboration for content quality assessment, optimized algorithms for more accurate user profiling, and enhanced interest exploration capabilities [34]. Additionally, natural language processing technologies can identify potentially misinformation-containing articles, and authority rankings can be elevated while demoting low-quality content to reduce misinformation visibility.

Second, learn from foreign methods for detecting visual misinformation and actively cooperate with third-party technical institutions and universities to explore deepfake identification. Short-video platforms like Douyin and Kuaishou have massive user bases, and visually disseminated misinformation is difficult to identify, requiring technological solutions for batch fact-checking. Chinese platforms remain relatively weak in this area. In August 2019, the "Zao" app attracted attention for enabling simple face-swapping operations, creating entertainment value while raising concerns about deepfakes and copyright infringement. Although later banned, the covert and low-cost nature of online operations necessitates both regulating deepfake technology use and enhancing capabilities to combat image and video deepfakes.

Third, platform-based media should establish deep cooperation with fact-checking organizations. Currently, independent fact-checking or rumor-debunking platforms have limited audiences and influence. Platform-based media can collaborate with these platforms using "labels" to enhance debunking effectiveness. Weibo' s @WeiboRumorDebunking and WeChat' s "WeChat

Rumor Debunking Assistant" send debunking information to users through regular push notifications, but only reach users who subscribe and click. Facebook' s labeling approach amplifies dissemination effects by attaching fact-checking results as labels to misinformation-containing posts, helping users quickly assess authenticity at a glance.

Furthermore, China should vigorously promote automated fact-checking technology R&D and implementation. Weibo and WeChat currently employ certain technical measures—Weibo monitors highly-retweeted information using technical means, conducting deep searches and expert verification when potential misinformation is detected; WeChat primarily uses technical interception, blocking content identified as misinformation by partners like People' s Daily Online, Guokr, and DXY during dissemination [35]. However, Chinese platform-based media should learn from foreign counterparts in cooperating with independent fact-checking organizations and developing automated fact-checking technologies, advancing automated fact-checking development through specialized collaborations and financial support to improve misinformation combat efficiency.

The intelligent media era will witness a technology-centered "offense-defense battle" in misinformation production and governance [36]. Technology can both facilitate and prevent misinformation. Deepfakes create reality-distorting videos indistinguishable from authentic content; malicious social bots spread across platforms, amplifying misinformation and misleading public opinion; misinformation achieves exponential growth through algorithmic recommendation mechanisms on platform-based media. This necessitates that platform-based media use technical means to identify misinformation in massive information flows and automatically generated/disseminated falsehoods, suppressing spread through labeling, interception, and deletion. However, technology has limitations—automated misinformation detection requires human journalistic professionalism and information literacy to operate effectively. Promoting collaborative participation among algorithmic technology, government, platforms, news media, universities, research institutions, nonprofit organizations, and users represents the inevitable trend in cyberspace governance.

References:

[1] Jerome Sun. Platisher: Another form of media convergence from Silicon Valley [EB/OL]. https://medium.com/@jeromesun_{66925}/%E5%B9%B3%E5%8F%B0%E5%9E%8B%E5%AA
%E6%9D%A5%E8%87%AA%E7%A1%85%E8%B0%B7%E7%9A%84%E5%8F%A6%E4%B8%80%E7%A7%8I
f8af5979e6ce.

[2] Jiao Jie. Platisher: A new type of converged media [J]. Western Journal, 2015(1): 20.

[3] Mikael Thalen. Eric Trump keeps falling for fake ballot hoaxes [EB/OL]. https://www.dailydot.com/debug/eric-trump-keeps-falling-for-fake-ballot-hoaxes/.

[4] Caldarelli G, De Nicola R, Del Vigna F, et al. The role of bot squads in the political propaganda on Twitter [J]. Communications Physics, 2020(1).

[5] Emily Saltz, Claire Leibowicz. Fact-Checks, Info Hubs, and Shadow-Bans: A Landscape Review of Misinformation Interventions [EB/OL]. https://www.partnershiponai.org/intervention-inventory/.

[6] Rachel Kraus. Facebook labeled 180 million posts as "false"since March. Election misinformation spread anyway [EB/OL]. https://sea.mashable.com/tech/13294/facebook-labeled-180-million-posts-as-false-since-march-election-misinformation-spread-anyway.

[7] TikTok. Taking action against COVID-19 vaccine misinformation [EB/OL]. https://newsroom.tiktok.com/en-gb/taking-action-against-covid-19-vaccine-misinformation.

[8] Facebook for Government, Politics and Advocacy. Understanding Facebook's Fact-Checking Program [EB/OL]. https://www.facebook.com/government/policies/fact-checking/misinformation-resources.

[9] Roshan Sumbaly, Mahalia Miller, Hardik Shah, et al. Using AI to detect COVID-19 misinformation and exploitative content [EB/OL]. https://ai.facebook.com/blog/using-ai-to-detect-covid-19-misinformation-and-exploitative-content/.

[10] Roshan Sumbaly, Mahalia Miller, Hardik Shah, et al. Using AI to detect COVID-19 misinformation and exploitative content [EB/OL]. https://ai.facebook.com/blog/using-ai-to-detect-covid-19-misinformation-and-exploitative-content/.

[11] Andrew Hutchinson. New Study Finds that Flagging False Reports on Facebook May Indeed Reduce their Distribution [EB/OL]. https://www.socialmediatoday.com/news/new-study-finds-that-flagging-false-reports-on-facebook-may-indeed-reduce-t/559968/.

[12] Ecker U, O' Reilly Z, Reid J S, et al. The Effectiveness of Short-Format Refutational Fact-Checks [J]. British Journal of Psychology, 2020(1): 36.

[13] Shannon Mullery. How the TikTok Algorithm Works in 2021 [EB/OL]. https://tinuiti.com/blog/paid-social/tiktok-algorithm/.

[14] Fang Shishi. News values behind algorithmic mechanisms—A study on the "Facebook bias gate" incident [J]. Shanghai Journalism Review, 2016(09): 41.

[15] Guy Rosen. An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19 [EB/OL]. https://about.fb.com/news/2020/04/covid-19-misinfo-update/.

[16] Statista. Most popular social networks worldwide as of January 2021, ranked by number of active users [EB/OL]. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

[17] Shannon Mullery. How the TikTok Algorithm Works in 2021 [EB/OL]. https://tinuiti.com/blog/paid-social/tiktok-algorithm/.

[18] Hazel Baker. Introducing the Reuters guide to Manipulated media, in association with the Facebook Journalism Project [EB/OL]. https://www.reuters.com/article/rpb-hazeldeepfakesblog/introducing-the-reuters-guide-to-manipulated-media-in-association-with-the-facebook-journalism-project-idUSKBN1YY14C.

[19] Viswanath Sivakumar, Albert Gordo, Manohar Paluri. Rosetta: Understanding text in images and videos with machine learning [EB/OL]. https://engineering.fb.com/2018/09/11/ai-research/rosetta-understanding-text-in-images-and-videos-with-machine-learning/.

[20] Jeremy Kahn. Facebook says it's made a big leap forward in detecting deepfakes [EB/OL]. https://fortune.com/2021/06/16/facebook-detecting-deepfakes-research-michigan-state/.

[21] Ferrara E, Varol O, Davis C, et al. The Rise of Social Bots [J]. Communications of the ACM, 2014(7): 96.

[22] Varol O, Ferrara E, Davis C A, et al. Online Human-Bot Interactions: Detection, Estimation, and Characterization [J]. 2017(1): 280.

[23] MarketingDive. Instagram may have 95M bot accounts, The Information reports [EB/OL]. https://www.marketingdive.com/news/instagram-may-have-95m-bot-accounts-the-information-reports/528141/.

[24] Stefan Wojcik. 5 things to know about bots on Twitter [EB/OL]. https://www.pewresearch.org/fact-tank/2018/04/09/5-things-to-know-about-bots-on-twitter/.

[25] Bovet A, Makse H A. Influence of fake news in Twitter during the 2016 US presidential election [J]. Nature Communications, 2019(7): 1.

[26] Bobby Allyn. Researchers: Nearly Half Of Accounts Tweeting About Coronavirus Are Likely Bots [EB/OL]. https://www.npr.org/sections/coronavirus-live-updates/2020/05/20/859814085/researchers-nearly-half-of-accounts-tweeting-about-coronavirus-are-likely-bots.

[27] Ferrara E, Varol O, Davis C, et al. The Rise of Social Bots [J]. Communications of the ACM, 2014(7): 100.

[28] Li Yangyang, Cao Yinhao, Yang Yingguang, et al. A survey of social network bot detection [J]. Journal of China Academy of Electronics and Information Technology, 2021(3): 214.

[29] DFRLab. #BotSpot: Twelve Ways to Spot a Bot [EB/OL]. https://medium.com/dfrlab/botspot-twelve-ways-to-spot-a-bot-aedc7d9c110c.

[30] Parag Agrawal. Twitter acquires Fabula AI to strengthen its machine learning expertise [EB/OL]. https://blog.twitter.com/en_{us}/topics/company/2019/Twitter-

acquires-Fabula-AI.

[31] Giorgia Guglielmi. The next-generation bots interfering with the US election [EB/OL]. https://www.nature.com/articles/d41586-020-03034-5#ref-CR2.

[32] Mei Song. Research on internet information governance from the perspective of national security [J]. Annual of Social Governance and Rule of Law, 2016(0): 101.

[33] Yi Qianliang. The "online gatekeeper" role of network platforms in content governance [J]. Youth Journalist, 2020(7): 24.

[34] Yu Guoming, Du Nannan. Value iteration of intelligent algorithmic distribution: "Boundary adjustment" and legitimacy enhancement—A case study of Toutiao' s four upgrades [J]. Shanghai Journalism Review, 2019(11): 19-20.

[35] Lu Shangqing. Research on the dissemination mechanism of social media rumors [D]. Jinan: Shandong University, 2016: 38, 40.

[36] Zhang Chao. Algorithmic governance of fake news on social platforms: Logic, limitations, and collaborative governance models [J]. Journalism and Mass Communication, 2019(11): 28.

Author Biographies: Shen Jinxia (1975-), female, from Minquan, Henan, Ph.D. in Journalism, Associate Researcher at the Institute of Internet Information, Communication University of China. Research interests: public opinion governance, internet content production and dissemination. Zhang Jiaying (1997-), female, from Nanyang, Henan, Master' s student at the Institute of Internet Information, Communication University of China.

(Executive Editor: Chen Xuguan)

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