Journal of Contemporary China

ISSN: 1067-0564 (Print) 1469-9400 (Online) Journal homepage: www.tandfonline.com/journals/cjcc20

'I Am Patriotic, but…': The Political Disclaimer as a Social Media Strategy in China

Tony Huiquan Zhang, Yingzhu Pu & Kecheng Fang

To cite this article: Tony Huiquan Zhang, Yingzhu Pu & Kecheng Fang (30 Aug 2025): 'I Am Patriotic, but…': The Political Disclaimer as a Social Media Strategy in China, Journal of Contemporary China, DOI: 10.1080/10670564.2025.2546903

To link to this article: https://doi.org/10.1080/10670564.2025.2546903

View supplementary material

Published online: 30 Aug 2025.

Submit your article to this journal

Article views: 193

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=cjcc20

JOURNAL OF CONTEMPORARY CHINA
https://doi.org/10.1080/10670564.2025.2546903

'I Am Patriotic, but…': The Political Disclaimer as a Social Media Strategy in China

Tony Huiquan Zhang [a], Yingzhu Pu [a] and Kecheng Fang [b]

[a]The University of Macau, Taipa, Macau SAR, China; [b]The Chinese University of Hong Kong. Hong Kong SAR. China

ABSTRACT

Chinese social media has become increasingly intolerant due to state control, platform regulation, and mob censorship. In reaction to this pressure, Chinese netizens use a strategy we term 'political disclaimer', whereby they make pro-regime or patriotic statements before voicing any criticism of the government or its policies. We argue Chinese social media users face both top-down and bottom-up pressures. Drawing on data collected from two major social media platforms in China during 2018–2022 and 66 in-depth interviews, influencers, and administrators, we found evidence of the widespread adoption of political disclaimers. Chinese netizens use them to achieve the following goals: (1) avoid censorship; (2) preclude critique; (3) legitimize requests; (4) attract sympathizers. However, our evidence suggests the effectiveness of political disclaimers is limited.

KEYWORDS

Political disclaimer; political communication; authoritarianism; nationalism; Social Media; China

Introduction

In April 2022, amidst the COVID-19 lockdown in Shanghai that triggered a significant backlash both domestically and internationally, an anonymous netizen posted the following on Zhihu, one of the largest social media platforms in China: 'Let me be clear, I support socialism—I just don't support the Zero-COVID policy'. The user begins the post by claiming to be pro-socialism, a 'politically correct' statement in China, thus attempting to pre-empt possible criticism of his stance against the government policy. The phenomenon of expressing support for the regime and the government before criticizing them has become common on China's social media in recent years. We conceptualize this intriguing phenomenon as a 'political disclaimer', referring to the practice whereby people make pro-regime and/or patriotic statements before expressing critical opinions to avoid negative consequences. In this article, we examine the trends and features of Chinese netizens' use of political disclaimers, explore the factors contributing to this strategy, and investigate their functions and potential effects.

Governments, particularly authoritarian ones, tend to strongly regulate online speech[1] and intervene in what can be spread in the public sphere. The Chinese government is particularly

CONTACT Yingzhu Pu pyzsociology@gmail.com Faculty of Social Sciences, University of Macau, Macau SAR, Taipa E21B-3001, China

[1]Robert Brym, Melissa Godbout, Andreas Hoffbauer, Gabe Menard, Tony H Zhang, 'Social Media in the 2011 Egyptian Uprising' [2014] 65(2) *The British Journal of Sociology* 266-292 <https://doi.org/10.1111/1468-4446.12080> accessed 15 January 2024; Min Jiang, 'Managing the Micro-Self: The Governmentality of Real Name Registration Policy in Chinese Micro Blogosphere' [2016] 19(2) *Information, Communication & Society* 203-220 <https://doi.org/10.1080/1369118X.2015.1060723> accessed 15 January, 2024; Elizabeth Stoycheff, and others, 'Privacy and the Panopticon: Online Mass Surveillance's Deterrence and Chilling Effects' [2019] 21(3) *New Media & Society* 602-619 <https://doi.org/10.1177/1461444818801317> accessed 15 January, 2024.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/10670564.2025.2546903.

© 2025 Crown Copyright

powerful in cyberspace control,[2] and major technology companies assist the regime in controlling the Internet.[3] In addition to the regulation by the state and social media platforms, pro-regime and patriotic users energetically monitor and combat perceived negative comments, viewing them as threats to social stability, regime legitimacy, and state security.[4] Scholars noticing this phenomenon have variously referred to it as 'cyber patrol,[5] 'participatory censorship',[6] and 'mob censorship'.[7] Regardless of the label, state, platform, and mob censorship have jointly created an unwelcoming environment for netizens wishing to voice their discontent publicly, making the use of political disclaimers an appealing social media strategy.

Political disclaimers are not unique to China, but part of a broader discursive strategy used to navigate discussions around sensitive topics in different contexts. In Western liberal democracies, individuals may employ disclaimers to preface potentially controversial views, especially if they contradict dominant ideological or social norms. For example, individuals may say, 'I am not a racist, but ... ' or 'I support women's rights, but ... ' when expressing views diverging from progressive values.[8] Similarly, moderate supporters of a controversial political figure, such as Donald Trump, may use qualifiers like 'I am not a Trump fanatic, but ... ' to signal a more nuanced or distanced position.[9] Individuals in democratic societies use these discursive strategies to navigate the boundaries of political correctness. Such pressure could be stronger in autocracies, with critics appearing in both the larger society (e.g. dominant views, social norms) and the state's official ideological narratives.[10]

[2]Jinjin Liu, Yingzhu Pu and Tony H Zhang, 'Critical Citizens and Discontented Citizens: Education, Perceptions of Fairness, and Support for Anti-COVID Measures in China' [2025] *Political Psychology* <https://doi.org/10.1111/pops.13079> accessed 21 July 2025.

[3]Siu-yau Lee, 'Surviving Online Censorship in China: Three Satirical Tactics and Their Impact' [2016] 228 *The China Quarterly* 1061-1080 <https://doi.org/10.1017/S0305741016001454> accessed January 15, 2024; Jiayin Lu and Yupei Zhao, 'Implicit and Explicit Control: Modeling the Effect of Internet Censorship on Political Protest in China' [2018] 12 *International Journal of Communication* 23 <https://ijoc.org/index.php/ijoc/article/view/8532> accessed 15 January, 2024; Luwei R Luqiu, 'The Cost of Humour: Political Satire on Social Media and Censorship in China' [2017] 13(2) *Global Media and Communication* 123-138 <https://doi.org/10.1177/1742766517704471> accessed 15 January, 2024; Tony H Zhang, Jianhua Xu and Jinjin Liu, 'How Do Toothless Tigers Bite? Extra-Institutional Governance and Internet Censorship by Local Governments in China' [2025] 261 *The China Quarterly* , 196-215. <https://doi.org/10.1017/S0305741024000602> accessed 21 July 2025.

[4]Rongbin Han, 'Defending the Authoritarian Regime Online: China's "Voluntary Fifty-Cent Army"' [2015] 224 *The China Quarterly* 1006-1025 <https://doi.org/10.1017/S0305741015001216> accessed 15 January, 2024; Rongbin Han, 'Withering Gongzhi: Cyber Criticism of Chinese Public Intellectuals' [2018] 12 *International Journal of Communication* 1966-1987 <https://ssrn.com/abstract=3151876> accessed 15 January 2024; Maria Repnikova and Kecheng Fang, 'Authoritarian Participatory Persuasion 2.0: Netizens as Thought Work Collaborators in China' [2018] 27(113) *Journal of Contemporary China* 763-779 <https://doi.org/10.1080/10670564.2018.1458063> accessed 15 January, 2024.

[5]Françoise Daucé, and others, 'From Citizen Investigators to Cyber Patrols: Volunteer Internet Regulation in Russia' [2019] 11(3) *Laboratorium: Russian Review of Social Research* 46-70 <https://doi.org/10.25285/2078-1938-2019-11-3-46-70> accessed 15 January, 2024.

[6]Zhifan Luo & Muyang Li, 'Participatory Censorship: How Online Fandom Community Facilitates Authoritarian Rule' [2024] 26(7) *New Media & Society* 42-36 <https://doi.org/10.1177/14614448221113923> accessed 21 July 2025.

[7]Priya Bhat and Kanchan Chadha, 'The Mob, the State and Harassment of Journalists via Twitter in India' [2023] 11(10) *Digital Journalism* 1788-1808 <https://doi.org/10.1080/21670811.2022.2134164> accessed 15 January, 2024; Jeffrey R Henrichsen and Matthew Shelton, 'Expanding the Analytical Boundaries of Mob Censorship: How Technology and Infrastructure Enable Novel Threats to Journalists and Strategies for Mitigation' [2023] 11(10) *Digital Journalism* 1848-1867 <https://doi.org/10.1080/21670811.2022.2112520> accessed 15 January, 2024; Silvio Waisbord, 'Mob Censorship: Online Harassment of US Journalists in Times of Digital Hate and Populism' [2020] 8(8) *Digital Journalism* 30-46 <https://doi.org/10.1080/21670811.2020.1818111> accessed 15 January, 2024.

[8]Waleed Aly and Robert M. Simpson, 'Political Correctness Gone Viral' in Chris Frost (ed), *Media Ethics, Free Speech, and the Requirements of Democracy* (Routledge 2014) 125–143.

[9]Jessica G Schafer, 'Donald Trump's "Political Incorrectness": Neoliberalism as Frontstage Racism on social media' [2017] 3(3) *Social Media + Society* <https://doi.org/10.1177/2056305117733226> accessed 15 January 2024.

[10]Rogier Creemers, 'Cyber China: Upgrading Propaganda, Public Opinion Work and Social Management for the Twenty-First Century' in King-wa Fu and Chung-hong Chan (eds) *Chinese Authoritarianism in the Information Age* (Routledge 2019) 15–30; Shaohua Guo, '"Occupying" the Internet: State Media and the Reinvention of Official Culture Online' [2018] 3(1) *Communication and the Public* 19 <https://doi.org/10.1177/2057047318755166> accessed 15 January 2024; Yingdan Lu and Jennifer Pan, 'The Pervasive Presence of Chinese Government Content on Douyin Trending Videos' [2022] 4(1) *Computational Communication Research* <https://doi.org/10.5117/CCR2022.2.002.LU> accessed 15 January 2024.

Both classical thinkers[11] and modern scholars[12] contend all political contexts have a dominant ideology. In theocracies, the state religion has the legitimacy and power to guide people's lives, actions, and speech. In autocracies, power holders may promote an official ideology to justify their rule (e.g. Marxism in most former Communist regimes). In democracies, the existence of a dominant ideology is arguable, yet there are certain ideological givens.[13] When individuals disagree with the dominant ideology, they may face negative consequences. Some are formal or institutional, such as legal consequences; others are informal and less evident, such as social sanctions and reputational damage. These consequences become pressures urging individuals to comply.[14]

Although political disclaimers are found around the world, China is unique in many respects. In some societies, the state strictly controls certain types of expressions; elsewhere, the public exerts pressure, as reflected in the 'cancel culture'. However, Chinese social media users face pressures from both sides. On the one hand, the state and social media platforms actively regulate expressions (top-down pressure). On the other hand, massive, passionate, nationalistic mobs voluntarily exert social control over regular users (bottom-up pressure).[15]

In this paper, we focus on the use of political disclaimers and argue for the uniqueness of the social media environment in China based on both quantitative and qualitative evidence. Our analysis of posts with political disclaimers shows that political disclaimers have become increasingly popular in China. Posts with political disclaimers are more critical and more likely to name government agencies and officials. Interestingly, these posts use fewer exclamation and question marks, implying the authors are cautious and attempt to express themselves in a controlled and less emotional manner. Our interviews with users and content moderators suggest Chinese social media users who are making political disclaimers are aware of China's media ecology and intentionally use disclaimers as a self-protective strategy. This explains why posts with disclaimers are more critical but less emotional.

We contribute to studies on political communication by probing the use of political disclaimers. We use China as a case study because of the power of the dominant political ideology and indoctrination[16] and because the government strictly controls online expression. Yet our findings have implications beyond China, as social pressures can be exerted by both the state/the platform and the mob in any location. Although we would expect to see users in other societies behave differently, they share the same basic logic: users need to adapt their criticism to meet the social demands of the world around them, including but not limited to cultural norms, religious taboos, political regulations, and economic incentives.

Literature Review: The Hostile Online Environment and Netizens' Strategies

The Chinese government maintains strict control of online content, with the goal of maintaining social stability and safeguarding the regime,[17] especially in the past decade.[18] Its methods

[11]Émile Durkheim, *The Elementary Forms of the Religious Life* (Courier Corporation 2012); Antonio Gramsci, *Prison Notebooks*, vol 2 (Columbia University Press 2011).

[12]Jonathan Hardy, *Critical Political Economy of the Media: An Introduction* (Routledge 2014).

[13]Louis Althusser, *On the Reproduction of Capitalism: Ideology and Ideological State Apparatuses* (Verso 2014).

[14]Lee (n 3); Lotus Ruan, Jeffrey Knockel and Masashi Crete-Nishihata, 'Information Control by Public Punishment: The Logic of Signalling Repression in China' [2021] 35(2) *China Information* 133-157 <https://doi.org/10.1177/0920203X20963010 > accessed 15 January 2024.

[15]Bhat and Chadha (n 7); Henrichsen, Jennifer R., and Martin Shelton, Expanding the analytical boundaries of mob censorship: How technology and infrastructure enable novel threats to journalists and strategies for mitigation [2023] 11(10) *Digital Journalism https://doi.org/10.1080/21670811.2022.2112520*; Waisbord (n 7); Silvio Waisbord, 'Mob Censorship Revisited: Questions, Findings, and Challenges' [2023] 11(10) *Digital Journalism* 1761-1768 <https://doi.org/10.1080/21670811.2023.2214905 > accessed 15 January 2024.

[16]Guo (n 10).

[17]Repnikova and Fang (n 4).

[18]Creemers (n 10); Zhang, Xu, and Liu [n 3].

include compulsory real-name registration, the Great Firewall preventing access to foreign websites, keyword detection and monitoring, automatic and artificial content removal, and manual moderation of content.[19] Netizens may find their critical comments have been hidden or deleted; worse yet, their social media accounts may be suspended or revoked. After years of this type of treatment, most Chinese netizens have learned to be silent or to voice with caution.

State pressure is not the sole source of the hostile online environment. Direct state control of online expressions can be costly; therefore, only issues with high priority will attract the state censorship machine—such as content exposing national leaders' scandals or calling for collective actions.[20] The state depends on local tech companies to handle other issues. China's large Internet population and highly developed virtual economy have nurtured local social media services (e.g. Weibo, TikTok, WeChat), and the country's policy barriers have shielded these service providers from competition with global equivalents (e.g. Twitter/X, Facebook, Instagram, WhatsApp). In return, these local companies are required to adhere to the government's regulations and policies and provide the necessary technical infrastructure and support for monitoring and censorship.[21] Tech companies who do not comply face penalties such as fines and the suspension or revocation of licenses.[22] As a result, all social media platforms in China comply with the state's expectation of censorship.

Compliance from social media platforms is not without hesitation, as it always comes at a cost. Since most social media platforms depend on user-generated content (UGC), strict censorship may hurt user experience, discourage UGC, and even drive users to alternatives. Yet companies need to obey to survive. For example, Weibo has established self-regulation and assists the government's propaganda goals by using automated algorithms for surveillance and censorship.[23] Another way for social media platforms to enforce censorship is to create a self-governing system and mobilize regular users—mainly nationalistic, pro-regime users—to collectively determine the fate of politically sensitive posts.[24] This 'participatory censorship'[25] has been adopted by many social media companies, including Zhihu, Tieba, and others (see Table 1)[26].

Mob censorship, namely the suppression or restriction of speech, expression, or information by regular citizens or web users, is another source of online pressure. In China, the issue most likely to attract mob censorship is nationalism. The recent rise of nationalist sentiments among Chinese youth has attracted scholarly attention.[27] Chinese young people have mobilized political campaigns against foreign media, foreign netizens, and domestic dissidents,[28] including the 'Anti-CNN' movement in 2008, anti-Japanese movements and sentiments in the 2010s, the 'DiBa Expedition' in 2016, and the anti-BBC campaign around the Xinjiang

[19]Gary King, Jennifer Pan and Margaret E Roberts, 'How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument' [2017] 111(3) *American Political Science Review* 484-501 <https://doi.org/10.1017/S0003055417000144> accessed 15 January 2024; Stoycheff, and others (n 1).

[20]King, Pan, and Roberts (n 19).

[21]Daucé, and others [n 5].

[22]Jiang (n 1); Yingdan Lu, and others, 'Decentralized Propaganda in the Era of Digital Media: The Massive Presence of the Chinese State on Douyin' [2025] *American Journal of Political Science* <https://doi.org/10.1111/ajps.12990> accessed 21 July 2025.

[23]Andy Zhao and Lily Hu, 'Unveiling Strategic Governance and User Dynamics in Weibo's Community-Driven Content Moderation System' [2025] *Information, Communication & Society* accessed 21 July 2025

[24]Yingjie Fan and Xu Xu, 'State-Endorsed Moral Outrage: Crowdsourced Coercion in the Era of Social Media' (unpublished manuscript, 2025).

[25]Luo and Li [n 6].

[26]See https://www.chinanews.com/fz/2013/10-22/5407770.shtml & https://money.ycwb.com/2019-06/14/content_30279073.htm; and '互联网陪审团'简史:从阿里大众评审到知乎仲裁官" Link: https://www.woshipm.com/it/2603043.html.

[27]Kecheng Fang and Maria Repnikova, 'Demystifying "Little Pink": The Creation and Evolution of a Gendered Label for Nationalistic Activists in China' [2018] 20(6) *New Media & Society* 2162-2185 <https://doi.org/10.1177/1461444817731923> accessed 15 January 2024; Han (n 4).

[28]Han [n 4]; Jennifer Pan, Xu Xu and Yiqing Xu, 'Disguised Repression: Targeting Opponents with Non-Political Crimes to Undermine Dissent' *Journal of Politics*, [2025] <https://doi.org/10.1086/734267> accessed 21 July 2025.

Table 1. List of social media platforms and year of implementation of self-governing systems

Social media platform	Self-governing system	year of implementation
Baidu Tieba 百度贴吧	Tieba Self-Governance Team 贴吧自律专员	2012
Sina Weibo 新浪微博	Weibo Community Council 微博社区委员会	2017
Dazhong Dianping 大众点评	Dazhong Juror 大众评审官	2017
WeChat 微信	Complaints Arbiters 投诉合议小组	2018
Zhihu 知乎	Zhihu Jury System 知乎众裁官	2019
Bilibili 哔哩哔哩	Discipline Board Member 风纪委员	2019
Xianyu 闲鱼	Xianyu Court System 闲鱼小法庭	2019
Douyin (TikTok in China) 抖音	Community Self-Governance Council 社区自律委员会	2020
Kuaishou 快手	Kuaishou Community Judge 快手社区评审员	2020
Meituan 美团	Meituan Juror 小美评审官	2021
Xiaohongshu (RedNote) 小红书	Community Supervisor 小红书社区监督员	2022

Data collected from public records, news reports on the Internet, and the official financial reports of social media companies.

controversies.[29] In recent years, such mobilizations have gone to the next level: regular users who are perceived as opposing state ideologies or policies are sought out and harassed.

Chinese netizens have developed various strategies to evade censorship and let themselves be heard. One strategy is to use homophones, puns, pseudonyms, and other keyword replacement techniques to avoid algorithmic and human censorship.[30] Another strategy is to use political satires, jokes, or *egao (*恶搞) to express political discontent.[31] In addition, Chinese netizens have gradually learned to read between the lines and extract useful information from official statements.[32] Similarly, they use what they have learned to express politically sensitive messages in a seemingly harmless way.

Our examination of political disclaimers builds on previous scholarship and adds new theoretical insights to the research on political expression and communication. The practices discussed above are largely evasive and indirect, trying to gain space for expression without crossing the 'red line' of ideological regulations. In contrast, online expression with a political disclaimer is more direct, straightforward, and unmediated. Users who make political disclaimers hope to openly comment on sensitive issues; they are aware of the unwelcoming environment, and they try to navigate it by signalling their political loyalty, or at least their compliance with ideological doctrines (e.g. the narratives of socialism or nationalism). As we go on to show, users of disclaimers are aware of the social media context and are therefore more carefully when being vocal.

[29]Peter Gries and Yi Wang, 'Proscribing the "Spiritually Japanese": Nationalist Indignation, Authoritarian Responsiveness, and Regime Legitimation in China Today' [2021] 245 *The China Quarterly* 122-141 <https://doi.org/10.1017/S0305741020000326 > accessed 15 January 2024; Han [n 4].

[30]Astrid Nordin, 'Un-Innovative Censorship, Innovative Resistance: The Internet, Forbidden Words and the Humorous Homonyms of Egao' in Astrid Nordin (ed), *Chinese Politics and International Relations* (Routledge 2014) 175–198; Weiming Ye and Luming Zhao, '"I Know It's Sensitive": Internet Censorship, Recoding, and the Sensitive Word Culture in China' [2023] 51 *Discourse, Context & Media* 100,666 <https://doi.org/10.1016/j.dcm.2022.100666 > accessed 15 January 2024.

[31]Luqiu (n 3); Guobin Yang and Min Jiang, 'The Networked Practice of Online Political Satire in China: Between Ritual and Resistance' [2015] 77(3) *International Communication Gazette* 215-231 <https://doi.org/10.1177/1748048514568757 > accessed 15 January 2024.

[32]Jiangnan Zhu and Chengli Wang, 'I Know What You Mean: Information Compensation in an Authoritarian Country' [2021] 26(3) *The International Journal of Press/Politics* 587-608 <https://doi.org/10.1177/1940161220963572 > accessed 15 January 2024.

Data and Methods

We explored the processes and mechanisms of political disclaimers using both quantitative and qualitative data. Our quantitative data were drawn from two social media platforms, Weibo and Zhihu. We chose Weibo and Zhihu because they are the primary social media platforms in China that welcome user-generated content and allow comments. These features encourage their users to speak up, yet they also enable mob censorship when controversies emerge, thus giving us the opportunity to observe political disclaimers and the reactions to them. We set the time frame for observation as 2018 to 2022; this period featured important political events in China and the world, including the global pandemic, trade wars, and geopolitical conflicts in Europe and the Middle East. Therefore, we expected that searching for political disclaimers generated during this period would yield rich findings.

We used web crawling with keywords and pattern searching to collect data. We collected social media posts under several highly political and controversial themes in China, including COVID-19 policy, the Sino-US trade war, foreign affairs and international relations, social inequality, sex and gender issues, and racial and ethnic issues, among others. A more detailed list of themes and sample keywords can be found in Table A1 in the Appendix. Since these themes are politically charged and controversial, they can provoke antagonism among social media users; we speculated individuals expressing these themes would more likely make political disclaimers in anticipation of criticism.

Before analysis, we filtered irrelevant or invalid search results (e.g. social media posts with too few characters, posts containing labels but no substantial content, etc.) and retained those with substantially long and meaningful content. This step yielded 2,576 politically related posts. Then, we manually cleaned and coded the posts based on whether they contained political disclaimers or not. Our final sample contained 1,354 pieces with political disclaimers and 1,222 pieces without disclaimers. Some exemplary disclaimers can be found in Table A2 in the Appendix.

Our qualitative data included 66 in-depth interviews with users, influencers, and platform administrators. The interviewees comprised 43 social media users and 12 influencers (number of followers/fans > 5,000) who had made political disclaimers. In our interviews, we asked about their experience in online expression on Chinese social media, their motivations to engage in political expression, their reasons for using political disclaimers, and their perceptions of the practice. We also interviewed 11 community administrators or content moderators hired by Weibo and Zhihu. For these administrators and moderators, we added a module asking how social media platforms expect them to work, how they censor politically controversial content, and their perceptions of political disclaimers. Our interview outline is available in Table A3 in the Appendix.

Results

Political Disclaimer: Trends and Content Features

> I love China; I love Huawei products too. But it is true that Ms. Meng Wanzhou holds so many foreign passports which looks suspicious. (Zhihu.com [QS-B-2021–0047])

This quotation is from a regular user on Zhihu. This user starts a critique of Huawei with 'I love China', which may sound like it comes out of nowhere and is quite bizarre for readers outside China and unfamiliar with the context of Chinese social media today. In recent years, Huawei—China's leading tech giant in telecommunications equipment, smart devices, and consumer electronics—has become a symbol of China's success in technological innovation and a national icon. Online users who favour Huawei products often identify themselves as patriotic Chinese buyers, and this nationalistic sentiment grew even stronger when the Sino-US trade war started to affect Huawei.[33] In December 2018, Huawei's CFO, Ms. Meng Wanzhou, was

[33]Wenna Zeng and Colin Sparks, 'Popular Nationalism: Global Times and the US—China Trade War' [2020] 82(1) *International Communication Gazette* 26-41 <https://doi.org/10.1177/1748048519880723> accessed 15 January 2024.

arrested in Canada on fraud charges. She was portrayed as an innocent victim of America's bullying behaviours during the Sino-US trade war, and her release and return to China in 2021 were celebrated nationwide.

Chinese netizens expressing negative views about Huawei or Ms. Meng—a heroine of China—seem to exercise caution, arguably because they face opposition from other Internet users. Many Huawei supporters view critics of Huawei products as being unpatriotic or having a blind preference for foreign goods. Others are nationalists who see Huawei as a national icon and source of pride, and they believe any criticism may be influenced by foreign companies or foreign ideologies.[34] Nationalistic views can be massive, organized, and pervasive, making it difficult for individuals to speak out without facing pressure or being labelled unpatriotic; simply stated, they use political disclaimers to avoid being targeted. The following example from Zhihu clearly uses political disclaimers, and the user explains why it is necessary:

> I am a 26-year-old patriot. I used to be a fan of Huawei. I initially chose Huawei as my first preference, but due to their unreasonable prices and excessive use of patriotic marketing, I ultimately decided to switch to another brand. While many Chinese citizens are supportive and excited about Meng Wanzhou, I personally do not share the same sentiment. However, I am hesitant to express my views openly as I fear online attacks from individuals who may label me as unpatriotic. (Weibo.com [QS-B-2020–0313])

Another theme around which we found disclaimers in our data was the global pandemic and anti-pandemic measures, especially during 2021–2022 when a Zero-COVID policy was strictly executed in China. The policy sparked a contentious debate on social media platforms[35] with users targeting the policy, its implementation, and the unintended consequences.[36] Despite the criticism, the government remained committed to the Zero-COVID policy until the end of 2022, when a wave of protests erupted nationwide against it. Before its dramatic termination, the policy enjoyed state endorsement and propaganda, making it immune to critiques, and a considerable number of people defended and justified it.[37] Those who criticized it risked being labelled political dissidents or foreign agents aiming to undermine China's anti-pandemic efforts, threaten social stability, and even topple the government. Given this chilling effect, political disclaimers often appear before critiques: '*I am not against the Zero-COVID policy, but* it has gone too far. The costs become unbearable for the economy and society' (Zhihu.com [QS-A-2022–0075]).

These examples suggest users think expressing a negative attitude towards Huawei or Zero-COVID policy may be interpreted as a deviation from national interests or official ideologies, so they use a political disclaimer to deflect criticism. Figure 1 shows the growing popularity of political disclaimers as an online strategy, particularly after 2020; both the number and the proportion of political disclaimers increased between 2018 and 2022.

We also looked at the tone of the posts. In general, political content on Chinese social media can be categorized as 'positive', 'neutral', and 'negative', based on users' attitudes towards the central government, local governments, establishments, political systems, and the overall society. We asked three research assistant coders to manually code the posts and ensured a satisfactory level of inter-coder reliability. Figure 2 displays the outcome. Posts with political disclaimers are clearly more negative and critical than those not using disclaimers.

We also explored whether social media posts mention certain government agencies or offices or name certain officials. It takes more courage to mention offices or individuals, as such posts are likely

[34]Yawei Chen and others, 'The Comments of Chinese "Zhihu" Netizens on the US Sanctions against Huawei: The Role of Anti-Western Centrism in Nationalist Narratives' [2022] 10(2) Asian Journal for Public Opinion Research 102-112 accessed 15 January 2024

[35]Yingdan Lu, Jennifer Pan and Yiqing Xu, *Public Sentiment on Chinese Social Media during the Emergence of COVID-19*, Twenty-First Century China Center Research Paper 4 (2021) <http://dx.doi.org/10.2139/ssrn.3757135 accessed 15 January 2024.

[36]Zhenyu Wang & Yuzhou Tao, 'Many Nationalisms, One Disaster: Categories, Attitudes, and Evolution of Chinese Nationalism on Social Media during the COVID-19 Pandemic' [2021] 26(3) *Journal of Chinese Political Science* 525-548 <https://doi.org/10.1007/s11366-021-09728-5 > accessed 15 January 2024.

[37]Suisheng Zhao, 'From Affirmative to Assertive Patriots: Nationalism in Xi Jinping's China' [2021] 44(4) *The Washington Quarterly* 141-161 <https://doi.org/10.1080/0163660X.2021.2018795 > accessed 15 January 2024.

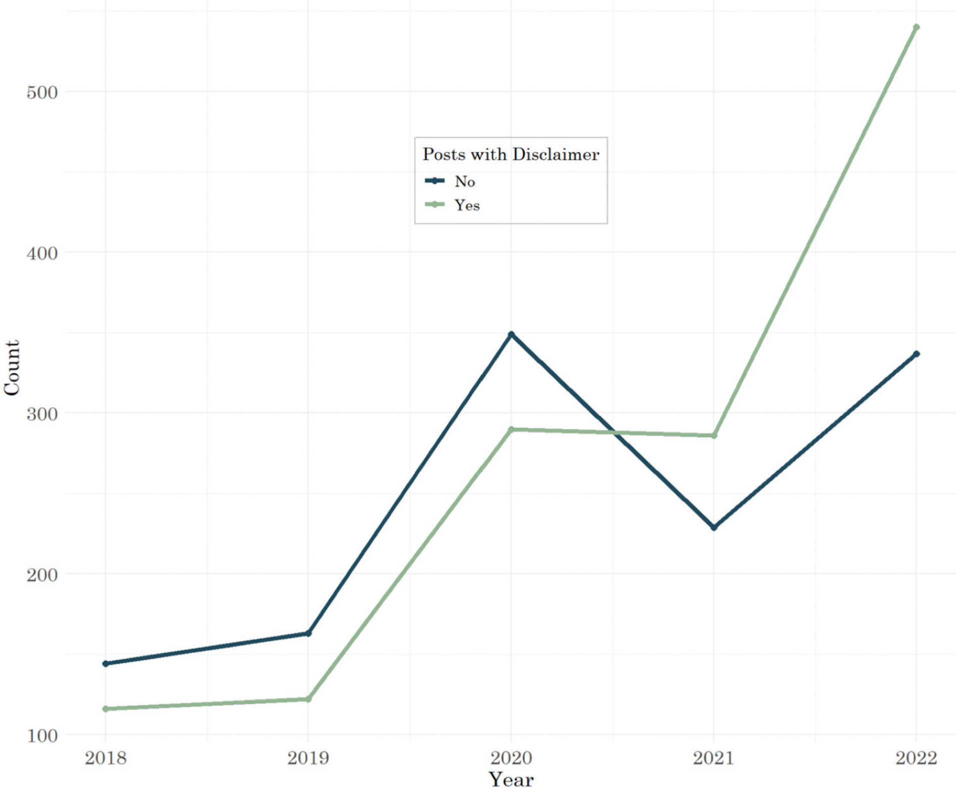

Figure 1. Increasing use of political disclaimers (2018–2022).

to trigger a response or even a backlash. We compared those mentioning offices/officials with those not naming anyone and plotted their tendency to make a political disclaimer (see Figure 3). Making a political disclaimer is a dominant strategy in both categories, yet those mentioning offices/officials tend to use disclaimers more than those not mentioning them. This suggests users criticizing governments and officials exhibit increased caution in response to perceived, or actual heightened risk.

Finally, we counted the number of exclamation and question marks in each social media post and calculated their density (the number of marks divided by the number of all characters). The results are displayed in Figure 4. Posts with political disclaimers are less emotional, with fewer exclamation and question marks. In other words, people using political disclaimers may be more aware of a possible backlash and thus exercise caution when voicing discontent online.

To sum up, our quantitative and descriptive results show political disclaimers became more popular and were more widely adopted during 2018–2022. Overall, posts with political disclaimers are more critical and more likely to mention names of government offices and officials. Despite being more critical and courageous, these posts are less emotional than those not using political disclaimers. Social media posts with political disclaimers usually have a substantial grievance or request and demand a policy change or ask for help, thus differing from general complaints without a clear target or agenda. Users mention names and target specific offices. Arguably, however, they are aware of the unwelcoming environment and were suitably cautious—this may explain why their posts are less emotional and use fewer exclamation and question marks. To validate this speculation

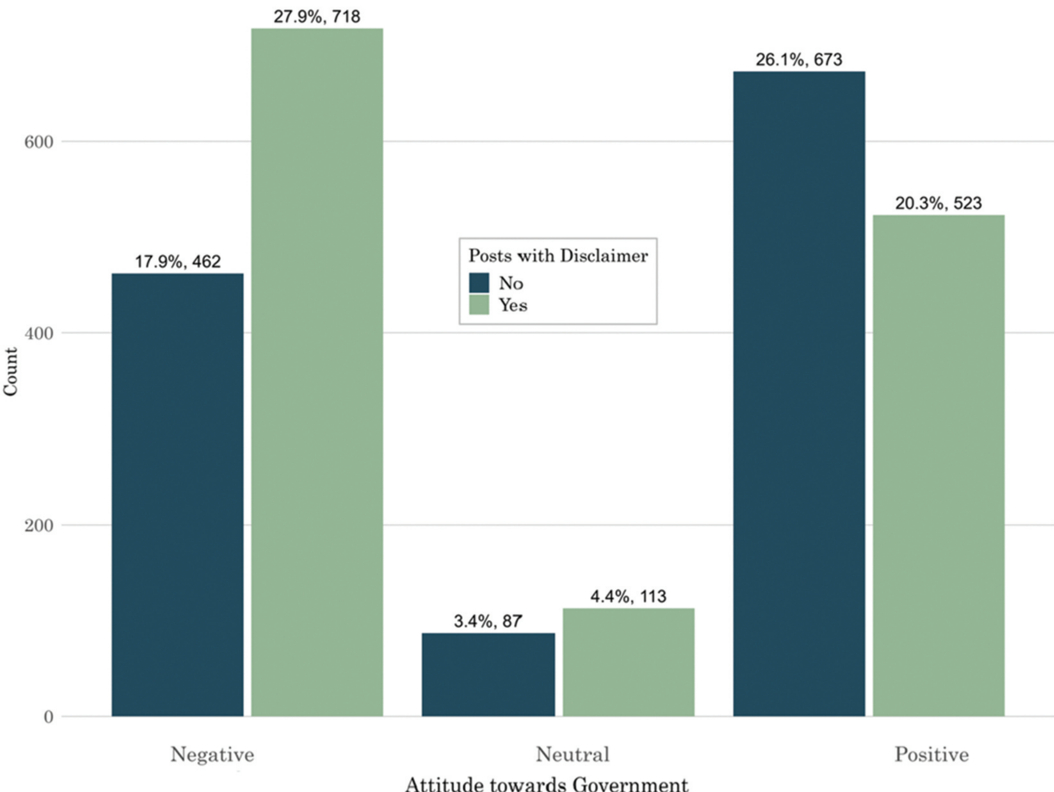

Figure 2. Overall division of posts: disclaimer vs. No disclaimer.

and understand why social media users feel the need to use political disclaimers, we turn to the results of our interviews.

Control and Patrol: Social Media Pressures from Above and Below

We organize the findings according to the two types of pressures users face, in line with the literature: top-down and bottom-up.

Top-Down Control: Roles of the State and the Platform

State control of Internet content is not new in China, with numerous studies highlighting practices such as propaganda and censorship.[38] However, endeavours to regulate the Internet have become increasingly subtle and implicit. For example, the state collaborates with celebrities and influencers in soft propaganda.[39] In addition, the state no longer directly confronts users and their content; instead, it pressures social media platforms to act on its behalf, by selectively rewarding and punishing companies for their performance in content control.[40]

[38]King, Pan, and Roberts [n 19]; Repnikova and Fang [n 4]; Guobin Yang and Wei Wang (eds), *Engaging social media in China: Platforms, Publics, and Production* (MSU Press 2021).

[39]Jian Xu and Ling Yang, 'Celebrity Public Relations in China: Power, Politics and Pop Propaganda' [2025] *The China Quarterly* <https://doi.org/10.1017/S0305741024001589> accessed 21 July 2025.

[40]Rebecca MacKinnon, 'China's Censorship 2.0: How Companies Censor Bloggers' [2009] *First Monday* <https://firstmonday.org/ojs/index.php/fm/article/view/2378> accessed 15 January 2024.

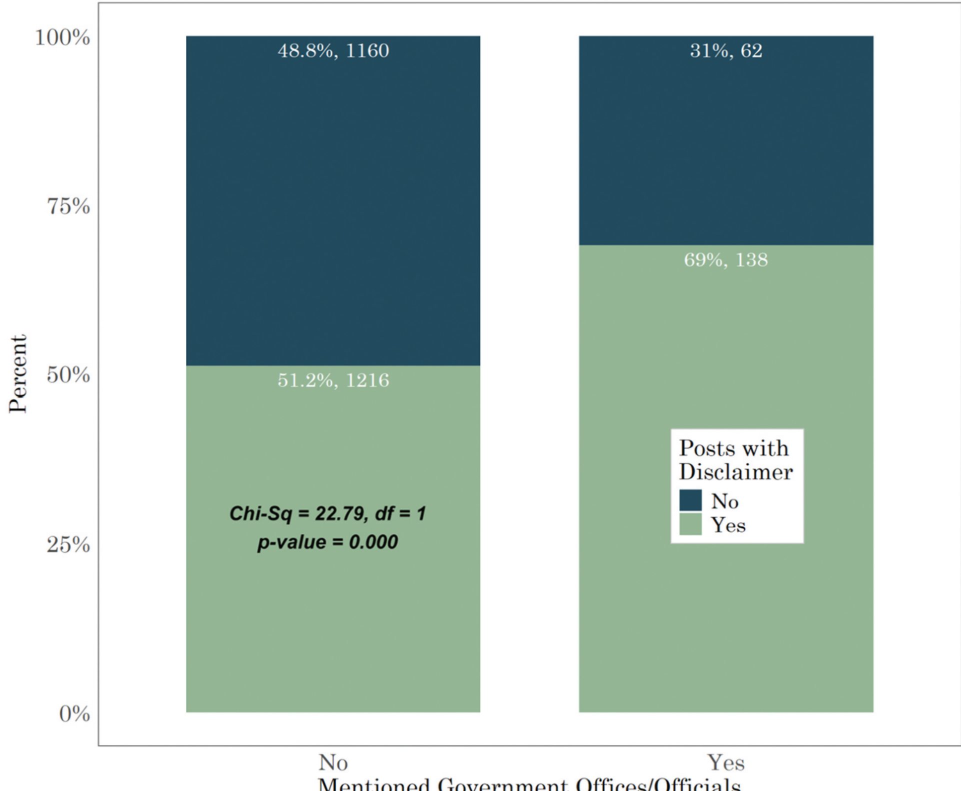

Figure 3. Posts mentioning government offices: disclaimer vs. No disclaimer.

Social media platforms intentionally regulate user communities based on state preferences and official ideological orientations. At times, regulation is based on direct orders from propaganda departments or cyberspace administration offices; at other times, it is based on platforms' interpretation of state policies, especially when they are unclear. As platform regulation may frustrate users or discourage content generation, many platforms have established community self-governance systems and created a buffer between platforms and users, so they do not need to directly face user dissatisfaction. In an interview, a Zhihu community administrator explained how companies recruit users to act as voluntary community moderators in daily operations:

> Both Zhihu and Weibo have established a system of jurors[41] consisting of community volunteers who are selected based on their alignment with the official attitude. These juries are given the right to review other users' content on the platform and providing suggestions to the platform, such as removal of the post, or suspension of the account. (Zhihu administrator [QS-X-2022–04])

As the comment suggests, Zhihu purposefully selects those aligning with official ideologies to be moderators or community jury members. In the long run, the atmosphere on Zhihu is becoming pro-regime and nationalistic, especially since 2017.[42] When an online environment shows a clear preference, users quickly realize what they should and should not do. A Zhihu administrator told us how users find where the 'red line' is:

[41]The official names for the platform juries are 微博社区管理员 (Weibo Community Administrators) and 知乎众裁官 (Zhihu Grand Juries) for Weibo and Zhihu, respectively.
[42]Altman Y Peng, 'Digital Nationalism versus Gender Politics in Post-Reform China: Gender-Issue Debates on Zhihu' [2022] 18(3) *Global Media and Communication* 281-299 <https://doi.org/10.1177/17427665221125537 > accessed 15 January 2024.

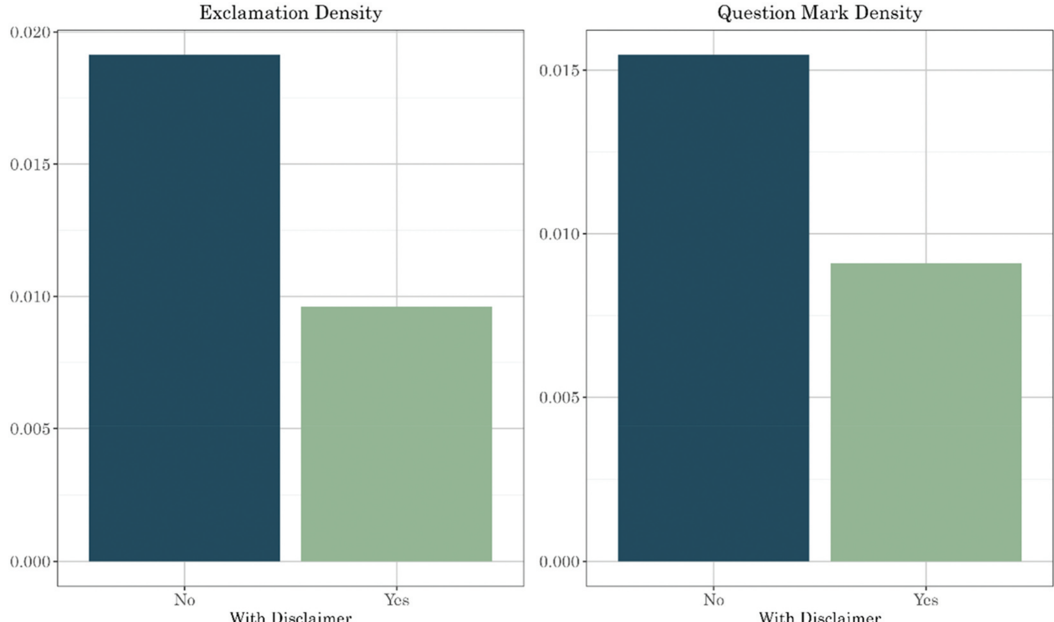

Figure 4. Exclamation and Question Mark Density: disclaimer vs. No disclaimer.

> Many videos related to the lockdown in Shanghai and articles criticizing the extreme anti-COVID measures in the city have disappeared quickly, indicating that such topics are not up for discussion. People soon learned where the boundary is; they learned to avoid certain topics to avoid repercussions. (Zhihu administrator [QS-X-2022–07])

For some complex and controversial themes, the state stance or platform preference is not always so clear, so users speculate about censorship and act accordingly. In such contexts, although the state remains unseen in online regulation and its expectations are ambiguous, its influence is omnipresent. One example is the so-called 'socialist recoding' trick used by gay writers on Chinese social media.[43] By replacing erotic content with 'Marxist-Leninist' keywords, the writers believe their 'recoding' will please the state apparatus and protect their content from censorship. The 'black box' of censorship criteria and the ambiguity of what the state wants makes platforms and users guess, speculate, and self-discipline.[44] The state power behind the social media platforms has a 'capillary power', which is more like daily guidance, moulding and discipline.[45]

A closely related phenomenon, 'algorithmic imaginaries', refers to users' imagination of how platform algorithms work in content moderation and traffic control. Users rarely understand the actual technical logic, but they often develop experiential knowledge and collective imaginaries[46] of how platform algorithms work. Wang[47] notes many Chinese social media users believe in and spread

[43]Ruichen Zhang, 'Negotiating Censorship through "Socialist Recoding" on the Chinese Internet: Nuances and Potentialities in a Contested Cyberspace' [2024] 27(1) *Information, Communication & Society* 126-142 <https://doi.org/10.1080/1369118X.2023.2179373> accessed 21 July 2025.

[44]Rachel E Stern and Jonathan Hassid, 'Amplifying Silence: Uncertainty and Control Parables in Contemporary China' [2012] 45 (10) *Comparative Political Studies* 1230-1254 <https://doi.org/10.1177/0010414011434295> accessed 15 January 2024; Ye and Zhao (n 30).

[45]J Cobbe, 'Algorithmic Censorship by Social Platforms: Power and Resistance' [2021] 34(4) *Philosophy & Technology* 739-766 <https://doi.org/10.1007/s13347-020-00429-0> accessed 15 January 2024.

[46]Lei Hao, 'Algorithmic Imaginary in Surveillance Technicity: The Renegotiation of Minzu Identity among China's Ethnic Minority Wanghongs on Douyin' [2024] *Global Media and China* <https://doi.org/10.1177/20594364241307478> accessed 21 July 2025; Ye and Zhao [n 30]; Zhu and Wang [n 32].

[47]Erika N Wang, 'Participatory Censorship with Illusory Empowerment: Algorithmic Folklore and Interpretive Labor beyond Fandom' [2024] 10(4) *Social Media + Society* <https://doi.org/10.1177/20563051241295800> accessed 21 July 2025.

'algorithmic folklore' or 'imaginations of censorship'. Although users may not have direct evidence to back up their speculations, their imagination of social pressures and algorithmic censorship is not unfounded, after years of experience on Chinese social media. As an interviewee in Wang's study put it, 'You accuse an enemy of 'an anti-communist stance' which may be dealt with more efficiently by platform administrators than if you accuse your enemy of obscene pornography".[48]

When users imagine the censorship machine will target political contents, they censor themselves accordingly. This is especially true when users rely on social media platforms for attention and income, as in the case of influencers.[49] A Zhihu-based content creator shared his experience of algorithmic uncertainty: "Sometimes I don't even know what triggered it. The post just doesn't get any views—barely two digits. When that happens, the post is dead'. The uncertainty results in more radical self-censorship. Consequently, the state has achieved an unprecedented level of social control without presence.

Patrol and Troll: Mob Censorship from the Bottom Up

> The people [who] disagree with me will create lots of Weibo accounts and then flood my page with comments. Even worse, they urge me to delete my posts by sending messages to my inbox, threatening my personal safety. (Weibo User [QS-X-2022-09])

In this comment, a Weibo user expresses fear of being harassed by the mob. Mob censorship (censorship implemented by regular users) makes the information environment restrictive and hostile. It involves large groups of people targeting individuals or organizations for their controversial (or perceived to be controversial) opinions or actions. It can be even more intimidating than state and platform censorship, as mobs are massive, spontaneous, unregulated and unpredictable. Mobs employ harassment, intimidation, or even doxing to silence their targets. This phenomenon was first noticed by Efrat (1990) long before the age of social media; with the explosion of social media, it became a trending phenomenon and has been captured by recent communication studies.[50]

Mob censorship is escalating online in China. Well-known cases include journalists and celebrities who dare to differ from official stances or ideologies, including Chai Jing[51] and Gao Xiaosong.[52] Chai Jing, a journalist who pays attention to human rights and environmental issues in China, released a documentary on China's air pollution in early 2015. Although the top environment official, Dr. Chen Jining (then Minister of Environmental Protection, currently Shanghai Chinese Communist Party Secretary) praised Chai and thanked her for her efforts, millions of Internet users attacked her for her elitist perspective or lack of patriotism. Mob censorship is not limited to harassing journalists such as Chai Jing; regular people often become the victims of users' voluntary participatory censorship[53] as well, when their comments are perceived as unpatriotic. One of our interviewees expressed her fear of personal harassment online, specifically the possibility of doxxing by the mob: 'I have posted my personal photos before; they could easily find out who I am and expose my information publicly—so I try my best to stay away from sensitive topics'.

Such pressure is more evident among influencers than regular users. As their income relies on their media output, Chinese influencers need to keep their accounts safe from deletion, suspension, or traffic limits imposed by platform algorithms. They do not want to irritate the state/platform or the

---

[48]Wang [n 48], 5.

[49]Elaine Jing Zhao, 'Wanghong: Liminal Chinese Creative Labor' in Stuart Cunningham and David Craig (eds), *Creator Culture: An Introduction to Global Social Media Entertainment* (New York University Press 2021) 211 <https://doi.org/10.18574/nyu/9781479890118.003.0015 > accessed 15 January 2024.

[50]Bhat and Chadha [n 7]; Henrichsen and Shelton [n 7]; Waisbord [n 7].

[51]Jie Zhang, *Chai Jing: The Power of Vulnerability* in *Female Celebrities in Contemporary Chinese Society* (2019) 39–61.

[52]For more detail, please see the work by Pan et al. (2025), and also news coverage such as 'Celebrities scrubbed from the Chinese internet', reported by Lianhe Zaobao, translated by Translated by Grace Chong, Candice Chan. Link: https://www.thinkchina.sg/society/celebrities-scrubbed-chinese-internet-victims-chinas-social-revolution

[53]Luo and Li [n 6].

regular users from whom they make profits. During an interview, an influencer with 140K followers on Weibo said:

> Just because I criticized the official behaviour on the incident of Pelosi's visit to Taiwan, a scolding war happened in my homepage, some users even sent private messages to scold me for being unpatriotic, and finally I had no choice but to delete this content. (Weibo User [QS-X-2022–10])

Mob censorship can reasonably be seen as a 'game-changer'. As an attack can happen at any time, it fundamentally changes the sense of security of those being attacked, making them more sensitive about their digital presence and footprint, especially those who need digital platforms to promote stories and to cultivate resources and reputation.[54] It is also noteworthy that the line between state-initiated smear campaigns and bottom-up attacks is increasingly blurred, contributing to a 'strategically ambivalent information operation'[55] to deter critical expression. Over time, this pressure has led to stronger, more elaborate, and more thoughtful disclaimers as the new norm on social media platforms in China. The length and complexity of disclaimers have increased as well. In fact, political disclaimers have become a survival tactic for netizens navigating a stressful online environment.

Intended Functions of Political Disclaimers

To better understand why and how people make political disclaimers, we created a 2-by-2 typology of disclaimers' intended functions based on the main target audience (state/platforms vs. other users) and the primary goal (preventative vs. proactive). Preventative goals mean people try to stop something from happening, such as avoiding state/platform censorship and precluding critiques from other users. Proactive goals mean people hope to get something by disclaiming, such as justifying their demands and attracting sympathy from bystanders. We summarize the four functions in Table 2. Note that this categorization is only an ideal type.

Although our typology categorizes political disclaimers based on their primary intended functions, in practice, these functions often coexist and reinforce one another. This is especially true in authoritarian contexts, where speaking candidly always carries political risks. Users must simultaneously express both loyalty and urgency to ensure their posts survive under state or platform censorship. In such an environment, avoiding censorship is the first goal. For example, during the COVID-19 lockdowns, users requesting government assistance had to first ensure their posts would not be censored or suppressed. The disclaimer was initially aimed at avoiding censorship (preventive), but it also served to legitimize demands and attract sympathy (proactive).

Our framework does not present four mutually exclusive categories. Rather, it highlights the strategic balancing acts users perform when navigating their expressive goals under conditions of political uncertainty. By adding a disclaimer at the beginning (occasionally in the middle or at the end), netizens hope to avoid triggering automated censorship algorithms or manual reviews by moderators. The disclaimer serves as a warning that the contents of the post may be sensitive or controversial, but the speaker's intention is benign. In the following example, Mr. K (a pseudonym), a Weibo influencer with nearly a million followers, says he hopes his post will survive the censorship:

Table 2. Intended functions of political disclaimers: a typology

Targets goals	State/Platforms	Other users
Preventive goals	Avoid censorship	Preclude critique
Proactive goals	Justify demands	Attract sympathizers

[54]Waisbord [n 15].

[55]Kecheng Fang, 'Wangbao (Cyberbullying) and Jubao (Reporting): Collaborative State—Society Online Influence Operations in China' [2024] 2(3) *Journal of Online Trust and Safety* <https://doi.org/10.54501/jots.v2i3.167 > accessed 21 July 2025.

> I support the measures for fighting the COVID-19 pandemic, but there are a few issues that need to be addressed. Firstly, it may not be safe for drivers to wear personal protective equipment while driving. Secondly, it seems unnecessary to require overseas travellers to undergo lengthy isolation periods for short visits. Lastly, it is unreasonable to mandate that citizens in areas with no COVID-19 cases must stay at home with a negative test result. (A post from Mr. K's Weibo timeline [QS-A-2021–0412])

Mr. K is obviously choosing his wording very carefully and expresses his overall positive attitude towards the Zero-COVID policy upfront. In our interview, we asked why he did so. Mr. K said a post criticizing a state policy usually gets deleted very quickly by the platform, but including a political disclaimer may help the post to survive. This survival may indicate that the platform has accepted a disclaimed post, or other users tolerate it and choose not to flag and report it to the platform, or both.

In addition, political disclaimers are often used to mitigate anticipated attacks. Netizens who engage in political debates usually have clear expectations of a possible backlash.[56] For example, the Russia-Ukraine war has been an enduring and controversial topic on Chinese social media since February 2022. Though never openly expressing it, the Chinese government's pro-Russia stance is clear, so pro-Russia users are often seen as siding with the government, while users with a pro-Ukraine tendency are seen as pro-US, pro-NATO, or even anti-China. The pro-regime vs. anti-regime dichotomy dominates public debates and is carefully fabricated and maintained by the state to deter dissenting voices.[57]

This environment is hostile to different opinions, especially those disagreeing with the regime. No wonder, then, that pro-Ukraine users on Chinese social media make political disclaimers when criticizing Putin. They do not want to be labelled foreign agents or 'American puppets'. Mr. P,[58] an anonymous Zhihu user, disagreed with China's pro-Russia stance, but only after he emphasized his status as a Communist Youth League member and advocated for China's safety in the international environment. Mr. P argued that allying with Russia post-2022 would be unwise for China, but said his argument was solely based on patriotism and national interest and nothing else. In the interview, Mr. P said he was worried that if he disagreed with official policy, he would be attacked by the nationalistic mob, so he centred on China's national interests and presented himself as a patriot.

The other type of goal is proactive—the user wants to get something by making a political disclaimer. Netizens sometimes make political disclaimers because they need governments and officials to help them. For example, netizens who were vocal about seeking medicine during the COVID-19 pandemic tried to express their needs without irritating officials. By saying 'I fully support the local government' or 'I love this country', users believed their demands would be met faster than 'undisclaimed' requests. In the following post, a Weibo user in Shandong Province demands an immediate end to the lock-down, but only after praising the government:

> *There's nobody who doesn't love their home. We support our government, and we love our homeland. We whole-heartedly understand the pandemic prevention efforts and support them*. However, right now we do not see any transparency in the decision-making process, and I do not understand why my city is locked down. This city has not been classified as a high-risk area, and there are no confirmed COVID infections at all. If that is true, we should resume a normal life and the lockdown must end; if there have been confirmed infections, we deserve to know that. (Weibo.com [QS-A-2022–0589])

Political disclaimers may be aimed at other users to gain their sympathy, thus reducing the number of hostile comments and possibly transforming into retweeting, attention, and sometimes

[56]Yinxian Zhang, Jiajun Liu and Ji-Rong Wen, 'Nationalism on Weibo: Towards a Multifaceted Understanding of Chinese Nationalism' [2018] 235 *The China Quarterly* 758-783 <https://doi.org/10.1017/S0305741018000863> accessed 15 January 2024.

[57]Fang and Repnikova [n 27]; Ruan, Knockel, and Crete-Nishihata [n 14].

[58]On Zhihu and Weibo, most users do not use real names, and gender information was unknown to us (until the online interviews). Yet most user profiles hint at 'gender' through user avatars or icons. This is usually set up by the users themselves. We relied on the user profile and used the speculated gender (or the confirmed gender information after the interviews) for the sake of convenience.

direct support (e.g. donations, volunteers). For example, a group of nurses posted a demand for help from the local government during the pandemic. Their demands were met after the post went viral; this would not have happened if it had been deleted, or the mob disliked it:

> We are a group of nurses in the city hospital who have been fighting the COVID-19 pandemic for months; we fully support the anti-COVID policy. Please do not harass us or curse us if you are reading this post. We are just a bunch of desperate nurses and could no longer bear the horrible working conditions. The entire hospital only gave us 50 gowns per week for the entire medical team. There was no proper sanitization or proper quarantine environment for patients – which got many of us infected. The hospital administrators do not respond to our requests for equipment and supplies. Now we call for more gowns, more protection, and we need to transport the infected medical workers to better hospitals and get some treatment! (Weibo.com [QS-A-2021–0617])

These cases suggest the predominance of preventive or proactive functions can hinge on the nature of the issue at stake. When users are personally affected by policies, such as lockdowns, medical shortages, or threats to livelihood, their posts typically reflect proactive goals aimed at prompting state or public responses. In such cases, the physical or material risks are immediate and tangible, and disclaimers serve as both protective devices and legitimizing frames for urgent demands. In contrast, for issues where users are not directly implicated, such as foreign policy debates or ideological conflicts, the function of disclaimers leans more heavily towards protecting discussion space online, particularly from nationalist mobs or politicized publics. Thus, while all disclaimers perform risk management, their dominant orientation—whether proactive or preventive—is shaped by the speaker's relationship to the issue, the immediacy of the perceived threat, and the locus of concern (material vs. symbolic).

To sum up, it seems Chinese social media users are fully aware of the social media environment; their political disclaimers are purposeful, and they serve both preventative and proactive functions. However, as we argue in the next section, political disclaimers can only get them so far. The state, the platforms, and the mob all have good reasons to oppose the speakers despite their use of political disclaimers.

Limitations of Political Disclaimers: Why the State, Platforms, and the Mob Reject Them

In our research, we noticed an increasing number of political disclaimers in online expressions, and our interviewees agreed that they felt the need to make disclaimers, thus suggesting political disclaimers are perceived as a useful means of self-protection. However, we also noticed that the effectiveness of political disclaimers has diminished. Why are the state, platforms, and the mob rejecting them?

First, an authoritarian regime may not buy what the political disclaimers are selling; when a post is deemed politically sensitive, radical, or even just inappropriate, the political disclaimer will not prevent censorship. Worse yet, in the age of algorithms and AI-assisted censorship, political disclaimers may draw attention to the post with political keywords such as 'I support socialism', 'I love the government', or 'I am not against the Zero-COVID policy'. In other words, political disclaimers may ironically lead to censorship.

Second, social media platforms have little incentive to tolerate a potentially controversial post, regardless of the presence or absence of a political disclaimer. As private companies under authoritarian rule, they are subject to government pressure and censorship mandates. In fact, platforms often play a harsher role in censorship than necessary to avoid negative repercussions,[59] such as being labelled unpatriotic or disloyal. The regime intentionally remains vague in its criteria for

[59]Guobin Yang, '(Un) civil society in digital China| Demobilizing the emotions of online activism in China: A civilizing process' [2018] 12 *International Journal of Communication* 21 <https://ijoc.org/index.php/ijoc/article/view/5313 > accessed 15 January 2024.

censorship, pushing platforms to engage in careful self-censorship.[60] In an interview, an administrator at Zhihu explained why she thought political disclaimers were useless:

> All content is subject to the scrutiny of both the algorithm and the human inspector. Before any web content becomes visible to users, we must ensure there are no political risks involved. So, it is important to note that even with political disclaimers, we need to go through the post content and see if it contains sensitive information. Saying that you are with good intentions is never enough.

In the interviews, several influencers complained that no matter how hard they try to avoid censorship, and no matter how long their disclaimer is, their posts can be deleted or at least face traffic reduction; in other words, their posts will not show up in other people's timelines or recommendations.

Members of the mob with a political agenda (e.g. Russian supporters, anti-feminists) also questioned the effectiveness of political disclaimers. An active feminist blogger, an influencer on both Zhihu and WeChat, said haters attacked her despite political disclaimers:

> Though I have been repeatedly saying that I am discussing gender issues only hoping for a better China, these people still firmly felt that I was provoking gender hostility, trying to sabotage Chinese society. Many speculated that I was an undercover agent hired by some Western feminist NGOs, or a spy sent by the CIA. (Zhihu & WeChat influencer [QS-X-2022–023])

The mob may refuse a disclaimer's validity for several reasons. Some online mobs are nationalistic and thus are suspicious of any message that appears to contradict the official Chinese Communist Party ideology or undermine regime legitimacy.[61] They may perceive any criticism or dissent as a threat to national unity, social stability, and territorial integrity[62] and therefore view a disclaimer as insincere and unconvincing. Even if netizens preface their comments with a political disclaimer, such as 'I am not trying to criticize the government. . .', they may still face a backlash from online mobs who see any discussion of these topics as unpatriotic.

Even dissident communities may reject those who use disclaimers. Dissidents may view those who claim to be loyal and patriotic as complicit in their regime's oppressive policies and practices and thus hypocritical or insincere. In their view, such disclaimers do not challenge or subvert the regime's propaganda and censorship.[63] Instead, they reinforce it by tacitly acknowledging the state's authority and official ideology, signalling the intolerant media environment and intimidating other users. Overall, the rejection of disclaimers by both the mob and dissident communities highlights the limitations of this media strategy in contemporary China.

Importantly, the failure of political disclaimers leads to divergent responses, which vary in their levels of popularity, platform dependency, and expressive goals. One content creator said, 'After building an account followed by 3,000 users over time, my account was repeatedly reported and eventually suspended for weeks'. To preserve her social media account (and her source of income), she said she deliberately reduced engagement with controversial topics and shifted to more neutral, informative content. As she put it, 'I no longer talk about politics much—just sharing and retweeting others' posts is safer'. In contrast, when users are not dependent on platforms, their reactions can be more resolute, as they have nothing at stake. A Zhihu user remarked: 'It is too hard to not violate any rules nowadays, when the "red line" is everywhere. So, I just don't censor myself anymore. They can suspend my account. No big deal'. (Zhihu User [QS-X-2023–01]) Users with more capital invested in a platform (such as influencers with more followers, content, and fame) tend to respond to backlash with caution and silence, while normal users may disregard the pressures from platforms and the mob. This highlights how situation and platform dependence affect the use of adaptive social media strategies in China.

[60] Ruan, Knockel and Crete-Nishihata [n 14].
[61] Han [n 4].
[62] Zhang, Xu, and Liu [n 3].
[63] Ruan, Knockel and Crete-Nishihata [n 14].

Discussion and Conclusion

In this study, we identified an emerging strategy on China's social media—the use of political disclaimers, characterized by an expression of political loyalty, nationalistic views, and pro-regime stances before criticism is voiced. Our data show political disclaimers are gaining popularity among Chinese netizens, and we argue this media strategy emerges in response to an increasingly intolerant social environment. This phenomenon appears in other societies as well, as a reaction to social pressures such as political correctness and the cancel culture. Yet in China, where top-down and bottom-up pressures coexist, social media users face more challenges, and the use of political disclaimers is more appealing. Our analysis of the political disclaimers we found on social media and the interviews with users, influencers, and administrators suggest the state, platforms, and other users exert pressure on those who would like to express criticism. Those making disclaimers hope to avoid censorship, preclude critique, legitimize requests, and attract sympathizers. Yet the effectiveness of political disclaimers seems limited; arguably, when more people use them, their efficacy dwindles.

This study makes several critical contributions to the literature on authoritarian politics and political communication. First, it highlights the agency of Chinese social media users in adapting to and navigating an increasingly intolerant online environment. Previous work on social media users in China and other authoritarian settings has noticed that netizens learn to express themselves with puns, homophones, political satire, and jokes.[64] Netizens use these strategies to evade censorship, gain popularity, identify allies, construct communities, and shield these communities from hostile outsiders. For example, Zhu & Wang[65] found netizens can infer and compensate for information deficiency through certain forms of communication, with successful communication enhancing feelings of solidarity. Similarly, Luqiu[66] found netizens use satire and humour not only to criticize the regime but also to construct their own community and identity awareness. We add to previous work by studying political disclaimers; our findings show users are aware of the intolerant social media environment and use disclaimers to signal their awareness of and compliance with the ideological norm. Unlike navigation strategies used by critical netizens to communicate among themselves, political disclaimers are intended to communicate with a larger audience base, trying to convey discontent and criticism to the public without irritating anyone. This social media strategy reflects both the wisdom of netizens and the changing social media context.

Second, the study suggests Chinese social media users are aware of what the state, social media platforms, and mobs hope to achieve in online social control. In our typology of functions of political disclaimers, users of disclaimers are addressing both top-down and bottom-up pressures. This adds to the literature showing how social media platforms assist the authoritarian state in web control and censorship. Previous work has noted how social media platforms boost official accounts, contents, or pro-regime materials with algorithms[67] and tolerate government-sponsored bots.[68] We found a subtler version of the platform's role—pressuring people to speak more cautiously. We believe the complex role of social media in balancing the need for community liveliness with platform survival in China merits further attention.[69]

Third, the study answers the call to understand China's authoritarian resilience[70] and evolving techniques of political control.[71] Our analysis of political disclaimers indicates censorship on Chinese social media is often not directly and explicitly implemented by the state; instead, platforms and mobs act on the state's behalf. In such a structure, the state can realize its goals without engaging in

[64]Luqiu [n 3]; Nordin [n 30]; Zhu and Wang [n 32].
[65]Zhu and Wang [n 32].
[66]Luqiu [n 3].
[67]Lu and Pan [n 10].
[68]Jiang [n 1].
[69]Xu and Yang [n 39].
[70]Joseph Fewsmith and Andrew J Nathan, 'Authoritarian Resilience Revisited: Joseph Fewsmith with Response from Andrew J Nathan' [2019] 28(116) *Journal of Contemporary China* 167-179 <https://doi.org/10.1080/10670564.2018.1511390> accessed 15 January 2024.
[71]Lu and Zhao (n 3).

daily monitoring and regulating, thus preserving both resources and legitimacy. After all, the pressures do not come from the state directly; voluntary control by platforms and mobs can prevent accusations of regime censorship and reduce dissident harassment. The state's implicit presence online[72] is a flexible yet effective way of maintaining online 'social stability'.

Fourth, our qualitative data point to the relationship between social media platforms and users. Regular users and influencers have different levels of platform dependence and differ in their use of political disclaimers—or more generally speaking, in their adaptative strategies in the social media environment. Influencers tend to be more cautious of and reactive to platform regulation and pressures, as they need the revenue and exposure; in contrast, regular users have more diverse reactions and are more reluctant to adapt. This resonates with broader discussions about how social media platform regulations and algorithms shape content production and influencers' identity construction processes.[73]

Finally, although we noted the rising popularity of political disclaimers, we also noticed their limited efficacy. This has important implications for freedom of speech in China and authoritarian societies in general. How will political disclaimers develop as a media strategy? They may become less common over time, as netizens increasingly recognize their constraints. Users may realize political disclaimers cannot adequately shield them from social backlash or censorship, leading them to self-censor.[74] However, the opposite may be true as well—we may see more people use political disclaimers and raise their intensity, despite their inefficacy. Like it or not, this kind of 'involution' (*nei juan* 内卷) among social media users may escalate as the social media environment deteriorates. Many users feel they are trapped in an awkward position; for example, one interviewee said to us:

> I think political disclaimers are of no use at all; they do not protect your post from deletion by the platform. However, you still have to say it. If you don't even say something at the beginning of your post, that's even worse. People will harass you right away.

This comment sums up the struggle of social media users in China if they want to express a political stance that deviates, even slightly, from the official line.

Admittedly, our work had some limitations. This was exploratory research discussing an emerging phenomenon, and we collected a limited amount of evidence over a short period (2018–2022). We anticipate there may be changes—disclaimers may or may not survive—and we call for future work to follow up on our findings. Moreover, although our selected themes and keywords (see Table A1 in the Appendix) are hot topics in China's public sphere, they may not be fully representative of social media discussions. Future work may benefit from employing a more comprehensive data collection strategy. The rise and the fall of disclaimers reflect changes in China's cyberspace. However, the implications of limited efficacy in political disclaimers are not all bad. Censorship can surprise users and trigger their curiosity in sensitive information.[75] When regular users all feel the need to disclaim (as our interviewees did) but find political disclaimers are less helpful, they may develop a better sense of their situation and come up with a new solution.

Acknowledgments

The authors would like to thank Juan Chen, Ze Hong, Jianhua Xu, Xinshu Zhao, and discussants at the International Chinese Sociological Association Annual Conference for their comments on earlier drafts of this paper. We thank the journal editor, Professor Suisheng Zhao and the anonymous reviewers for their constructive feedback. We appreciate the excellent research assistance provided by Wenyuan Fan, Zhixian Chen, Jinjin Liu, Yanghuayue Shi, Guyu Sun, Yourong Yao, and Yuan Zhuang.

[72]Zhang, Xu, and Liu [n 3].
[73]Hao [n 46]; Zhao [n 49].
[74]Waisbord [n 7].
[75]William R Hobbs and Margaret E Roberts, 'How Sudden Censorship Can Increase Access to Information' [2018] 112(3) *American Political Science Review* 621-636 <https://doi.org/10.1017/S0003055418000084> accessed 15 January 2024.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The study was funded by Multi-Year Research Grants [Grant Numbers: MYRG2022-00085-FSS & MYRG-GRG2024-00036-FSS] provided by the Faculty of Social Sciences, University of Macau.

ORCID

Tony Huiquan Zhang http://orcid.org/0000-0002-3587-5910
Yingzhu Pu http://orcid.org/0009-0004-9715-3085
Kecheng Fang http://orcid.org/0000-0003-1674-1543

CRediT Authors

Tony Huiquan Zhang: Conceptualization, Project administration, Research design, Writing—first draft and revision; Yingzhu Pu: Methodology, Data collection and analysis, Result visualization; Kecheng Fang: Conceptualization, Research design, Writing—revision & final editing. The authors share equal contributions to this article, and their names are listed in a reverse alphabetical order.

Data Availability Statement

Quantitative data (anonymised version) can be requested from the authors; qualitative materials are kept confidential to protect participants.

Research Ethics and Consent

The research was conducted under the guidelines and protocols of the University of Macau's Research Ethics Review System (RERS). We gained informed consent from all interviewees and protected the privacy and confidentiality of participants. The interview data were anonymised, and no information is identifiable.