JOURNAL Of

LANGUAGE, MEDIA & SOCIETY

Vol. 1 No. 1 | Spring 2026

Carter & Co. PUBLISHING

JOURNAL Of LANGUAGE, MEDIA & SOCIETY Vol. 1 No. 1 | Spring 2026 Carter & Co. PUBLISHING

JOURNAL Of

LANGUAGE, MEDIA & SOCIETY

Vol. 1 No. 1 | Spring 2026

Carter & Co.
PUBLISHING

Copyright © 2026 Carter & Co. Publishing
All rights reserved.
No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the publisher, except in the case of brief quotations used in scholarly review or academic work.

Journal Information
eISSN: 3071-1673
Print ISSN: Pending

Publisher
Carter & Co. Publishing
Website: www.carterandcopublishing.com
Journal Website: www.journaloflms.org

Peer Review Statement
All articles published in this journal are subject to peer review. The journal follows a double-blind review process to ensure academic rigor and integrity.

DOI Information
Articles in this journal are assigned Digital Object Identifiers (DOIs) to ensure persistent access and citation. DOI registration is managed through Crossref.

Disclaimer
The views expressed in the articles are those of the authors and do not necessarily reflect the views of the publisher or the editorial board.

Production Information
Printed in the United States of America

CONTENTS

EDITORIAL BOARD

Managing Editor
Yu Xiang, Carter & Co. Publishing, USA

Reviews Editor
Jinpu Wang, Metropolitan State University, USA

Editorial Board Members
Huseyin Zeyd Koytak, University of Mississippi, USA
Jianlin Chen, Shanghai International Studies University, China
Lijing Gao, Missouri University, USA
M. Aynal Haque, Montana Technological University, USA
Nabeel Siddiqui, Susquehanna University, USA
Nasiba Norova, Metropolitan State University, USA
Xin Zhao, Bournemouth University, UK
Zhifeng Kang, Fudan University, China

EDITORIAL NOTE

The Journal of Language, Media and Society (JLMS) is founded as an interdisciplinary platform that seeks to move beyond conventional disciplinary boundaries in the study of language, media, and society. As digital infrastructures, algorithmic systems, and artificial intelligence reshape the conditions of knowledge production and communication, existing disciplinary frameworks are often insufficient to capture these transformations. JLMS brings together diverse perspectives and methodological approaches to explore how meaning, mediation, and social relations are reconfigured in contemporary contexts.

This inaugural issue reflects the journal's commitment to interdisciplinary inquiry by presenting research that engages with a wide range of themes, including digital discourse, global media dynamics, and the sociocultural implications of technological mediation. Rather than aligning with a single disciplinary tradition, the journal encourages dialogue across fields and supports work that combines theoretical innovation with empirical analysis.

JLMS is committed to maintaining high academic standards through a double-blind peer review process and the guidance of an international editorial board. The journal aims to provide a space for both established scholars and emerging researchers to contribute to critical conversations at the intersections of language, media, and society. As Managing Editor, I would like to express my sincere gratitude to the authors, reviewers, and editorial board members whose support has made this inaugural issue possible. We warmly welcome future submissions and look forward to the continued development of JLMS as a space for rigorous, collaborative, and forward-looking scholarship.

Yu Xiang
Minneapolis, MN
April.22.2026

Journal of Language, Media and Society
Vol. 1, No. 1 https://doi.org/10.66056/jlms.v1i1.001

Carter & Co. PUBLISHING

SOCIAL MEDIA-MEDIATED STRATEGIES OF ANTI-RACISM FOR THE ASIAN COMMUNITY: A SYSTEMATIC LITERATURE REVIEW

XIN ZHAO

Urgent action is needed to combat the rise in anti-Asian hate exacerbated by the COVID-19 pandemic. This systematic literature review examines previous studies on social media-mediated anti-racism strategies, given the increasing scholarly interest in the role of social media in addressing social justice issues and the need to empower both Asian and non-Asian communities in anti-racism efforts. The paper reviewed 38 peer-reviewed studies, categorizing them based on key attributes such as publication outlets, geographic focus, and methodological approaches. It also reviewed the identified anti-racism strategies in the papers, as well as their interrelationships with agents, effectiveness, and outcomes. The review also documented the associated challenges. Based on the findings, this review proposes four key directions for future research: (a) expanding the scope of strategies through diverse scholarly perspectives, (b) deepening understanding of these strategies across different national, socio-cultural, and platform-specific contexts, (c) identifying patterns of effectiveness by triangulating findings from multiple methodological approaches, and (d) systematically examining the challenges of leveraging social media for social justice initiatives.

KEYWORDS: anti-racism; Asian; social media; systematic literature review

Introduction

Racial discrimination against Asians, particularly East and Southeast Asians (ESEA), has been pervasive and deeply ingrained in Anglo-American societies, manifesting in various forms such as stereotyping (Chou & Feagin, 2016) and exclusion (E. Lee, 2016). This issue was both exacerbated and brought to the forefront during the COVID-19 pandemic. Research has established a strong link between racism and negative mental health outcomes, including depression, anxiety, and psychological distress, among the global Asian community (D. L. Lee & Ahn, 2011). During the pandemic, racial discrimination further contributed to adverse consequences for the ESEA community, including diminished well-being (Cheah et al., 2020; S. Lee & Waters, 2021) and financial hardship (Huang, Krupenkin, Rothschild, & Cunningham, 2023). Addressing racism against the Asian community is therefore an urgent priority, requiring effective and sustained interventions.

During the pandemic, the Asian community actively utilized social media platforms to collectively combat racism. This marks a significant historical moment, as it may be the first time the community has mobilized on such a large scale to publicly challenge racial discrimination. This shift contrasts with the traditional mindset of earlier generations, which often emphasized

© 2026 Carter & Co.

avoiding conflict. For instance, reflecting offline collective actions against anti-Asian violence under the slogan "Stop Asian Hate," the hashtag #StopAsianHate has been used in 337.7K posts on the short-video platform TikTok (as of February 2025).

Existing scholarship on the relationship between social media and racism against the Asian community has primarily focused on two key areas: first, social media-mediated racism and its negative impact on the Asian community; and second, the underlying factors influencing engagement in social media-driven anti-racism efforts. More recently, scholars have begun examining the specific anti-racism strategies facilitated by different social media platforms in various contexts. For instance, emerging studies such as Xinyu Zhao and Abidin (2023) and J. J. Lee and Lee (2023) explore how the ESEA community has leveraged the affordances of the short-video platform TikTok to collectively address racism through various strategies, including raising awareness and fostering pan-Asian solidarity.

As Keum and Volpe (2023) astutely observe, research on online coping strategies for addressing racism remains limited, and the impact and effectiveness of these strategies are still unclear. This systematic literature review (SLR) aims to provide a comprehensive overview of existing studies on identified strategies, assess their effectiveness where applicable, and highlight future research directions for leveraging social media more effectively in the pursuit of social and racial justice. Theoretically, this study builds on ongoing academic discussions regarding the role of social media in activism and social justice movements. It also provides empirical insights into anti-racism efforts. Rather than perpetuating a narrative that frames racialized communities as powerless victims who passively endure racism (Corneau & Stergiopoulos, 2012; Ellefsen, Banafsheh, & Sandberg, 2022), this paper emphasizes their agency and autonomy in actively utilizing available tools—social media, in this case—to develop coping mechanisms (Corneau & Stergiopoulos, 2012, p. 277) and resistance strategies (Ellefsen et al., 2022). Recognizing this agency is crucial for informing policymakers, social media platform designers, and activist organizations in the design of targeted, social media-driven initiatives that further empower individuals and civil society in the fight against racism.

Social media as a hotbed of racism

Media and communication scholarship has long examined the prevalence and patterns of racist discourse against the Asian community on social media, including its production, circulation, and impact on both victims and society at large, particularly in the context of the COVID-19 pandemic.

A growing body of research has documented the widespread racial discrimination against the Asian community on social media during the COVID-19 pandemic, manifesting in various forms such as harassment, exclusion, threats, hostility, and the spread of misinformation (Shi et al., 2022; Shin, Wang, & Song, 2023; Tong, Stoycheff, & Mitra, 2022). Moreover, Uyheng, Bellutta and Carley (2022) found that bots played a significant role in amplifying hate speech in online discussions about racism during the pandemic.

Previous studies have consistently demonstrated the harmful impact of social media-mediated racism against the Asian community, affecting both community members and those outside the community. During the COVID-19 pandemic, exposure to racism on social media was linked to heightened concerns about real-world discrimination among young Australian Asians, which, in turn, contributed to negative emotions and lower life satisfaction (Shin et al., 2023). Similarly, in the U.S., reliance on social media for COVID-19–related news was positively associated with greater concerns about future discrimination among Asian community members (Yu, Pan, Yang, & Tsai, 2020). These findings align with research conducted outside the pandemic context. For instance, Lee-Won, Lee, Song, and Borghetti (2017) found that, compared to nonracist messages, microblogged racist messages elicited stronger feelings of anger and shame among Asian users. Furthermore, social media consumption also influenced out-group perceptions. Willnat, Shi, and De Coninck (2023) found that higher consumption of COVID-19-related news on social media among White Americans correlated with increased anti-Asian stigmatization, which in turn reinforced the perception that Asian immigrants were less deserving of entry into the U.S.

Anti-racism on social media

If the widespread online racial discrimination facilitated by social media represents the *yin*, then the simultaneous resistance to racism, such as anti-racism discourses and various forms of online activism, embodies the *yang*.

Social media can serve as a powerful platform for addressing a range of social justice issues, including medical disenfranchisement (Xin Zhao, Feigenbaum, & Demirkol Tønnesen, 2024), gender equality (Chaif & Finneman, 2024), climate justice (Hannouch & Milstein, 2025), to name a few. Thanks to their interactive infrastructure, these platforms facilitate raising awareness, collaboratively constructing and expanding related knowledge, critiquing underlying structural problems, and mobilizing calls to actions (Xin Zhao et al., 2024).

Specifically in relation to the goals of anti-racism, the facilitative role of social media has been particularly evident in the Black Lives Matter (BLM) movement. Social media-enabled discussions of the movement helped amplify marginalized voices (Nartey, 2022); foster internal connections, garner resources from outsiders and lay movement members, build coalition with other groups, and promote preferred narratives of the movement (Mundt, Ross, & Burnett, 2018); establish and expand public engagement through sympathy-based identification (Edrington, 2022); and motivate young students to participate in offline political activism (Clark, 2016). Notably, the momentum generated by the BLM movement on social media also predicted increased coverage of police brutality by mainstream news media, which could potentially channel anti-racism demands towards policymakers (Freelon, McIlwain, & Clark, 2018).

As to anti-racism for the Asian community, a growing body of research has examined the factors influencing online coping strategies in response to racism against the Asian community. These studies explore various determinants, including the roles of mobilizing activities and intrinsic and extrinsic motivation in predicting intention for online civic engagement in anti-Asian violence activism (Kang, 2023), the relationship between COVID-19-related racial discrimination (e.g. blame and assault) and civic engagement on social media (Park et al., 2024), and the impact of problem-focused and emotion-focused coping strategies on social media activism (Tao, Li, Lee, & He, 2024).

Research on online coping strategies for addressing racism is limited, and it remains unclear whether coping online with racism provided satisfactory coping experiences and allowed individuals to obtain effective online social support (Keum & Volpe, 2023).

Given the growing scholarly interest in the role of social media platforms in addressing

social justice issues, particularly in combating racism, this SLR aims to provide an overview of social media-mediated anti-racism strategies for the Asian community. Driven by this aim, this paper will answer the following research questions:

RQ1: In which journals and years were the studies published?

RQ2: What was the geographic focus (country), geographic focus (region), social media platform, context, and methodological approach that the studies focused on or used?

RQ3: What social media-mediated anti-racism strategies were identified, along with their agents, effectiveness, and outcomes?

RQ4: What issues were identified in leveraging social media to combat racism against the Asian community?

Methods

Data collection

The literature included in this SLR must meet four criteria: (1) the studied communication is mediated by social media, (2) it aims to address racism, (3) it is contextualized within racism-related events or activities, and (4) it pertains to the Asian community. To conduct the SLR, this paper adapted the protocols and steps used by Melchior and Oliveira (2022) and Lough and McIntyre (2023). This process involved (a) selecting search keywords, (b) applying these keywords to major scholarly databases, (c) scanning identified articles to determine inclusion or exclusion, and (d) conducting full-text readings of selected articles for further inclusion or exclusion decisions, and in the meantime, snowballing these articles.

First, the database Communication Source, available through the author's affiliation, was used to identify key terms related to relevant existing studies. As the most comprehensive full-text research database for communication studies, Communication Source includes 654 active full-text communication journals, making it a suitable resource for targeted keyword identification. Moreover, its automatic keyword suggestion feature enhances the comprehensiveness of the selected terms. After conducting trial searches, the following search string was formulated:

asian AND (racism or discrimination or prejudice or racial bias or race or stereotypes or racial inequality or anti-racism or antiracism or antiracists or anti-racist) AND social media

Second, the search string was applied to two databases, Communication Source and Web of Science, retrieving 52 and 611 articles, respectively. The full texts of located articles were then downloaded and stored them in Zotero, a free, open-source reference management tool that facilitates collaboration and automatically detects duplicates. To expand the dataset, an additional search using Google Scholar was conducted. The author screened titles for relevance (e.g. prioritizing those mentioning anti-racism rather than racism alone) and then keyword searched whether the article mentioned "social media" in the main text. The search was stopped at page 30, where saturation was reached, yielding 12 additional articles.

Third, the articles were scanned by examining their titles, abstracts, and main texts. This review included empirical studies rather than literature reviews or editorials, considering both journal articles and PhD theses, as both undergo rigorous peer review. Only articles written in English were included. This step also helped exclude irrelevant studies, such as those on anti-racism strategies that are not mediated by social media. However, this review retained

articles that, despite not explicitly mentioning social media-mediated anti-racism in their titles (e.g. studies only mentioning the examination of online racism discourses in their titles), still addressed the topic in their main text. After this screening, 69 articles remained.

Fourth, full-text reading allowed for a more detailed assessment of the articles' eligibility. This process led to the exclusion of studies in which Asian participants constituted only a small portion of the sample or where their perspectives were minimally represented. Notably, this review included articles that examined both effective and ineffective social media-mediated anti-racism strategies (e.g. general social media use linked to perceived discrimination), as well as those that did not explicitly measure effectiveness, as all fall within the scope of our SLR. This step left this study 31 articles. During this step, a snowballing approach was also employed, identifying seven additional articles by reviewing the references of selected studies and screening publisher-recommended papers. The author continued sampling new relevant publications until March 2025. As a result, the final sample comprised 38 articles.

Following the PRISMA flow diagram (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009), Table 1 below reported the research design.

Table 1. Data collection record.

Step A	Deciding search key terms		
Step B	Communication Source	Web of Science	Google Scholar
	52	611	12
Step C	Scanning		
	69		
Step D	Full-text reading	Snowballing	
	31	7	
Total	38		

Data analysis

Guided by the research questions, this review coded the articles based on the following aspects: journal title, year of publication, geographic focus (country), geographic focus (region), social media platforms, context, methodological approach, identified strategy, its agent (i.e. who applied the strategy), effectiveness, outcome, and associated issue.

The classification of "identified strategy" was informed by existing frameworks in intersecting fields, including models for conceptualizing coping strategies for racism (Brondolo, Brady, Pencille, Beatty, & Contrada, 2009), general computer-mediated coping activities (Hanasono & Yang, 2016), general social media coping activities (C. Yang & Tsai, 2023), and anti-racism strategies specific to the Asian community identified in the sampled articles. For example, this review adopted "cathartic expressions" from Abidin and Zeng (2020) as an overarching category encompassing various emotional expressions, such as anger, as identified by Brondolo et al. (2009). The categories of "associated issue" were derived from the sampled papers. Table 2 below presents the categories of "identified strategy" and "associated issue." For the code "effectiveness," this review only coded the quantitative studies that explicitly examined the effectiveness of specific strategies and left all other studies unclassified. The review did not categorize the code "outcome" but documented the specific findings in each study.

Table 2. Categories of "identified strategy" and "associated issue."

Code	Category	Operational definition
Identified strategy	General social media use	Using social media in a general sense (Cai, Ahmed, Ibasco, & Chib, 2024)
	Consumptive social media use	Following or reading social media content (Ahmed, Chen, Jaidka, Hooi, & Chib, 2021; Chen, Sun, & Tao, 2024)
	Expressive social media use	Engaging with social media by publishing, sharing, or commenting (Ahmed et al., 2021; Chen et al., 2024)
	General coping online with racism	Coping with racism in online settings (Keum & Volpe, 2023)
	Exchanging experiences	Exchanging lived or vicarious racism-related experiences to amplify the reach and visibility of anti-Asian racism (Abidin & Zeng, 2020)
	Cathartic expression	Expressing strong emotions associated with personal or vicarious experiences of racism (e.g. anger, frustration, sadness, disappointment, laughter, humor, or satire) (Abidin & Zeng, 2020; J. J. Lee & Lee, 2023)
	Self-presenting racial/ethnic identity	Publicly showcasing the subjective sense of group membership, focusing on shared history, values, or common heritage (Brondolo et al., 2009)
	Exchanging social support (to victims and/or non-aggressors)	Providing or receiving support (emotional, esteem, network, informational, or tangible supports) to address racism-related issues (Abidin & Zeng, 2020; Hanasono & Yang, 2016)
	Seeking social support	Asking for support to address racism-related issues (Brondolo et al., 2009)
	Assertive coping (toward aggressors)	Expressive communication about the incident, such as directly questioning the aggressor's behaviors, making an official complaint, demanding an apology, clarifying the intent of the act, or asserting what was inappropriate without resorting to insults (X. Wang, Wu, & Rajtmajer, 2023; F. Yang & Hanasono, 2021)
Identified strategy	Social media activism	Calling for collective action to address racism-related issues at the societal or structural level (Chon, 2023; Chon & Park, 2020; Jun, Kim, & Woo, 2024)
Associated issue	Counterproductive discourses	Invalidated, discredited, or insensitive discourses in response to anti-Asian racism (Abidin & Zeng, 2020)
	Negative impact on mental health	Mental health adversely affected when exposed to anti-racism related content on social media (Atkin, Ahn, Yi, & Li, 2024; C. Yang & Tsai, 2023)
	Ineffective platform functions	The inability of social media platforms to support or facilitate anti-racism efforts (Hanasono & Yang, 2016; Odağ & Moskovits, 2024; Parker & Song, 2006; Xinyu Zhao & Abidin, 2023)
	Fear of repercussions	Fear of unintended consequences when coping online (Odağ & Moskovits, 2024)

Coding

The author designed and conducted the data collection and analysis. To ensure coding reliability, they coded the articles twice, with a one-month interval between each round. An independent reviewer then reviewed the coding. We reached consensus on the coding for all quantitative studies. After further discussion, we refined the coding for "outcome" in studies using qualitative and mixed methods, adding more nuance to the coding.

Findings

Table 3 and Figure 1 below answered RQ1 that askes in what journals and when the studies were published.

Journal title

Table 3 shows that the *Asian Journal of Communication* publishes the highest number of papers (N = 7) on social media-mediated anti-racism strategies for the Asian community, followed by *Social Media + Society* (N = 4). This finding is unsurprising, as the focus of these papers aligns closely with the aims and scope of both journals: the former specializing in communication issues with an Asian perspective and the latter examining social media within social and cultural contexts. While journals in fields such as health, race, crime, psychology, linguistics, and regional media and communication also published relevant studies, they did so in much smaller numbers. Given that addressing anti-Asian racism through social media requires interdisciplinary collaboration, these results highlight the urgent need for continued research beyond media and communication studies. Furthermore, identifying nuanced strategies across different national and cultural contexts remains essential, as racial discrimination against the Asian community and the corresponding resistance efforts are deeply shaped by regional historical backgrounds.

Table 3. Journals where the papers were published.

Journal title	No. of papers
Asian Journal of Communication	7
Social Media + Society	4
Journal of Medical Internet Research	2
Frontiers in Public Health	1
Frontiers in Communication	1
Media International Australia	1
New Media & Society	1
Race and Justice	1
Journal of Counseling Psychology	1
Media, Culture & Society	1
Crime & Delinquency	1

Journal title	No. of papers
Telematics and Informatics	1
Journal of Current Issues & Research in Advertising	1
Canadian Psychology/Psychologie canadienne	1
Ethnic and Racial Studies	1
The Sociological Review	1
Popular Communication	1
Proceedings of the 15th ACM Web Science Conference 2023	1
Heliyon	1
Cyberpsychology, Behavior, and Social Networking	1
Howard Journal of Communications	1
Communication Quarterly	1
Multilingua	1
European Societies	1
Ethnicity & Health	1
International Journal of Environmental Research and Public Health	1
Asian American Journal of Psychology	1
Political Communication	1

Year of publication

Figure 1 illustrates a noticeable increase in publications on this topic from 2022 onward. The year 2022 marked the third year of the COVID-19 pandemic, which served as the contextual backdrop for 23 papers in the total sample (see "Contexts" below). This upward trend aligns with the typical lifecycle of peer-reviewed publications, which often take several years from research to publication. The number of publications peaked in 2023 (N = 11) and remained steady in 2024 (N = 10). This trend suggests a sustained academic interest in the topic, largely driven by the unprecedented public health crisis and its societal implications.

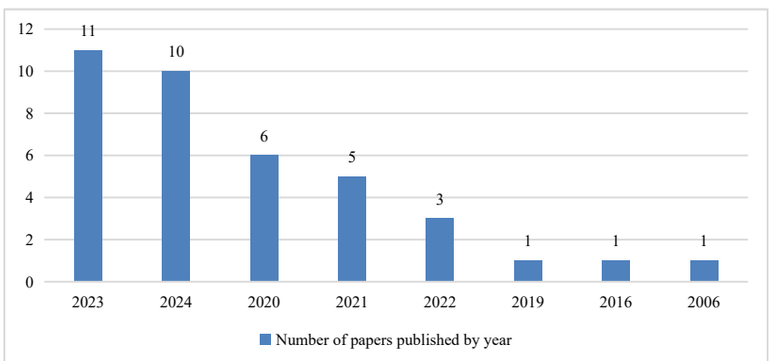


Figure 1. Number of papers published by year.

RQ2 enquired the geographic focus (country), geographic focus (region), social media platform, context, and methodological approach that the studies focused on or used.

Geographic focus

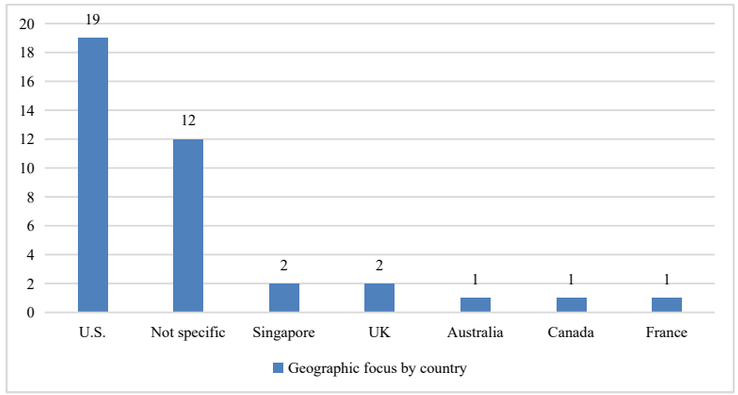


Figure 2. The geographic focus (by country) of the papers.

As shown in Figure 2, half of the sampled papers (N = 19) focused on the U.S. This finding reflects the significant role of the U.S. in the development and growth of the Stop Asian Hate movement, as well as its transition from offline to online spaces. Among the 12 papers not contextualized within a specific country, 11 did not specify a regional focus, while one was set in Europe. This is understandable, as it is often difficult to pinpoint the geographic locations of social media users when sampling online content. Only five other countries have been the subject of scholarly attention, each receiving limited focus. Given that anti-Asian hate is a global issue, and social media platforms have facilitated resistance efforts, there is an urgent need to broaden the scholarly exploration of social media-mediated anti-racism strategies for the Asian community to encompass more geographic regions. The fight against racism requires collective global efforts.

Social media platform

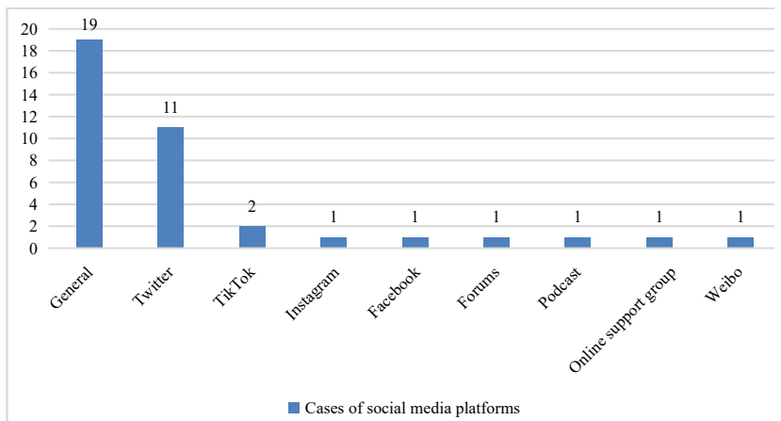


Figure 3. The social media platforms that the papers focused on.

Figure 3 shows that half of the sampled papers (*N* = 19) did not specify the exact social media platform(s) they focused on, instead examining social media use in a general sense. The emphasis on Twitter (currently known as X) highlights the platform's significance as a digital space for activism under the #StopAsianHate hashtag. Other social media platforms received only limited scholarly attention. As demonstrated in J. J. Lee and Lee's (2023) study, social media users creatively utilized platform features to construct anti-racism discourses and spaces. This finding underscores the need for more research into the diversity, nuances, and characteristics of anti-racism strategies across different social media platforms.

Contexts

23 papers were contextualized within the COVID-19 pandemic, and six focused on the Atlanta spa shooting, with three addressing both events. 12 papers did not specify a particular context. This result is not surprising, as there has been a surge in online resistance discourses in response to both online and offline racism incidents targeting the Asian/ESEA community, which were amplified by the pandemic. However, it also highlights the lack of academic attention to these strategies prior to the pandemic. As indicated in Figure 1 above, only three articles were published before 2020, when the pandemic began. This gap does not imply the absence of online coping strategies outside the pandemic context, nor does it diminish the need for further research in this area.

Methodological approach

19 studies employed quantitative research methods, 15 used qualitative approaches, and four utilized mixed methods. Among the quantitative studies, strategies such as consumptive social media use, expressive social media use, and social media activism were frequently quantified, with online surveys being the predominant method (e.g. Ahmed et al., 2021; Chen et al., 2024; Cho, Li, Cannon, Lopez, & Song, 2021; Jun et al., 2024). These quantifiable insights enable a deeper examination of the interrelationships among social media-mediated anti-racism strategies, their agents, effectiveness, and outcomes (see Figure 5 below for details).

Qualitative and mixed-methods studies provided nuanced perspectives on these strategies, documenting real-life cases and examples. Additionally, they captured strategies that were rarely or not at all addressed in quantitative studies. For instance, self-presenting racial/ethnic identity and seeking social support were exclusively explored in qualitative studies. Similarly, cathartic expression was identified in ten qualitative studies but appeared in only one quantitative study.

RQ3 askes the identified social media-mediated anti-racism strategy, as well as their agent, effectiveness, and outcome.

Social media-mediated anti-racism strategy

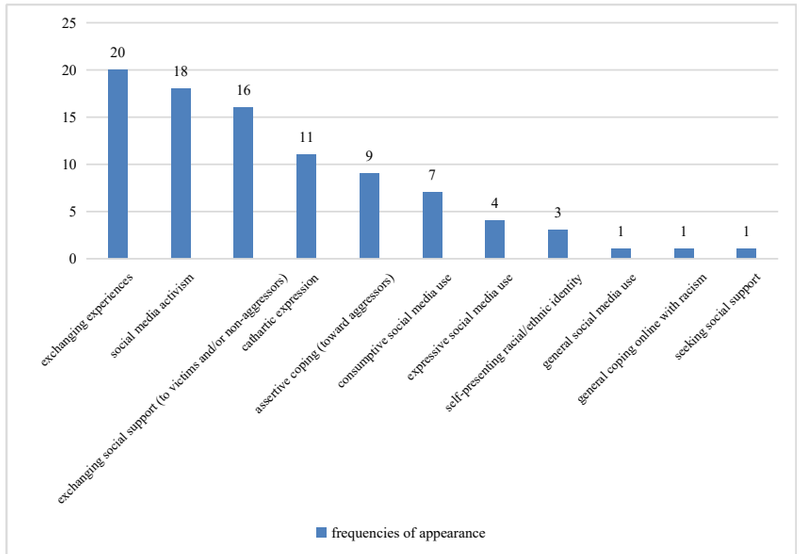


Figure 4. Frequencies of appearance of the identified social media-mediated anti-racism strategies in the papers.

As illustrated in Figure 4, the most frequently identified strategies include exchanging racism-related experiences (*N* = 20), social media activism (*N* = 18), exchanging social support (to victims and/or non-aggressors) (*N* = 16), and cathartic expression (*N* = 11). In contrast, other strategies, such as assertive coping (toward aggressors), self-presenting racial/ethnic identity, and seeking social support, were identified in less than ten papers. This finding aligns with previous research highlighting the facilitative role of social media platforms in sharing experiences of marginalization, building support communities, and mobilizing collective action for social justice (e.g. Xin Zhao, Feigenbaum, & Demirkol Tønnesen, 2024; Xin Zhao, Feigenbaum, & McDavitt, 2022). The reasons behind the limited focus on these other strategies remain unclear—whether due to a lack of scholarly attention or their relative rarity in real-life cases. However, the review of the challenges associated with social media-mediated anti-racism strategies for the Asian community (discussed below) may offer insights into their limited presence.

Agent, effectiveness, and outcome of the strategies

Of the reviewed papers, 24 focused on the Asian community as agents in implementing various social media-mediated anti-racism strategies, ten examined the general public, two explored non-Asian communities (e.g. individuals identifying as Caucasian), one analyzed business accounts, and one investigated Black and Asian activist organizations. These findings suggest a relatively balanced scholarly focus on different groups involved in anti-racism efforts for the Asian community.

Among the 19 quantitative studies, 14 explicitly examined the effectiveness of various strategies. Figure 5 below illustrated their research findings on the social media-mediated anti-

racism strategies, their agents, effectiveness, and outcomes.

Among the strategies analyzed, expressive social media use was consistently associated with positive outcomes across both Asian and non-Asian communities. Two studies found that social media activism among Asian community members was positively linked to an increased intention to participate in other forms of activism. These findings are reasonable, as both types of social media engagement suggest a high level of digital literacy and active participation, which may contribute to intrapersonal resilience and civic engagement.

Figure 5. Research findings of 14 quantitative studies that explicitly examined the effectiveness of the strategies.

The effects of consumptive social media use, however, produced inconsistent patterns across the three groups of agents. The patterns were contradictory among the studies targeting the general public and community members, respectively. Two studies found that social media consumption among non-community members predicted negative outcomes.

Other strategies and their relationships with agent, effectiveness, and outcomes were each examined in only one study. Notably, the positive impact of exchanging social support on social media users' well-being, as identified by F. Yang and Hanasono (2021), aligns with findings from other settings. For instance, Manohar and Kline (2024) found in an experiment that high person-centered emotional support (i.e. legitimizing and validating the recipient) and racial identity affirmation (i.e. ascribing value to the recipient's racial identity) were perceived as more effective in enhancing collective self-esteem (e.g. feeling good about belonging to one's racial/ethnic group) than lower-quality versions of these support types.

Interestingly, for Asian community members, none of the three strategies, i.e. general social media use, general coping online with racism, and exchanging experiences, were found to predict positive outcomes, despite the fact that, among the three strategies, exchanging experiences is the most frequently identified one among the sampled papers. This highlights the need for further research not only on *what* the social media-mediated anti-racism strategies are but also on *how* they impact Asian communities, non-Asian communities, and society as a whole (Keum & Volpe, 2023).

Among the studies employing qualitative and mixed research methods that touched upon the outcomes of the strategies, one notable pattern emerges. Research focusing on the strategy of exchanging experiences has consistently documented its positive impact on raising public awareness of anti-Asian racism and strengthening community cohesion not only within the Asian community (see Abidin & Zeng, 2020; Atkin et al., 2024; Cao, Lee, Sun, & De Gagne, 2022; S. Wang et al., 2021; Xinyu Zhao & Abidin, 2023; Zhu, 2020) but also in society at large (see Criss et al., 2023). This finding presents an interesting contrast to the quantitative study by Pan, Yang, Tsai, and Dong (2021), which identified a positive correlation between exchanging racism-related information on social media and increased depression during the COVID-19 pandemic. This contrast underscores the complexities of assessing the impact of anti-racism strategies, as they influence not only the internal states of community, as well as non-community, members (e.g. emotions, attitudes, thoughts, and perceptions) but also the broader network of relationships both within the community and between the community and non-community members. It is important to recognize that there may not be a one-size-fits-all strategy to address the diverse needs of all stakeholders in the anti-racism effort.

None of the studies explicitly examined the impact of self-presenting racial/ethnic identity on either group of agents. Rivas-Drake, Pinetta, Juang, and Agi (2022) proposed that youths' understanding of their ethnic-racial identities and their link to those of others can foster productive intergroup relations and, consequently, collective well-being. Given the popularity of social media among younger generations and the widespread practice of expressing ethnic and racial identity online, such as presenting the 'Asian + White' multiraciality on TikTok (King-O'Riain, 2022), it is crucial to investigate whether such self-presentation facilitates anti-racism.

RQ4 questions the issues identified in leveraging social media to combat racism against the Asian community.

Issues

Seven studies mentioned the inefficiency of platform functions in effectively supporting anti-racism efforts for the Asian community. Identified issues included the platform's inability to facilitate and manage in-depth civil discussions (Abidin & Zeng, 2020; Kuo & Jackson, 2024), the dominance of platform algorithms and filter bubbles over resistance discourses (Odağ & Moskovits, 2024), unresponsive platform assistive services (Odağ & Moskovits, 2024), challenges in integrating online discussions into decision-making processes (Parker & Song, 2006), limited engagement between in-groups and out-groups (Y.-J. Lee, Haley, & Shang, 2024), and low conversion rates from awareness to action (Xinyu Zhao & Abidin, 2023).

Five studies touched upon counterproductive discourses on social media related to resisting racial discrimination against the Asian community. These included inter-Asian and inter-minority tensions, as well as comparisons of racism-related suffering (Abidin & Zeng, 2020), counter-speeches that inadvertently reinforced racist stereotypes (X. Wang et al., 2023), and responses that failed to provide sufficient emotional support (Hanasono & Yang, 2016).

Two studies documented the negative mental health impacts of engaging with social media for anti-racism efforts, including stress and feelings of hopelessness (Atkin et al., 2024; C. Yang & Tsai, 2023). One study presented the community members' fear of potential repercussions from relevant social media engagement (Odağ & Moskovits, 2024).

These issues may explain the limited application of certain social media-mediated anti-racism strategies, as shown in Figure 4. Strategies such as assertive coping (toward aggressors), self-presenting racial/ethnic identity, and seeking social support may attract extreme hate speech while failing to garner sufficient support from both in-group and out-group members.

The issues identified in the sampled papers extend beyond anti-racism. They can also provide a reference to the challenges associated with social media-mediated efforts in addressing other social justice issues, such as disparities in medical diagnoses and treatment (Xin Zhao et al., 2024, 2022) and unequal access to health information (Xin Zhao & Xiang, 2023).

Discussion

Overall, the sampled papers were published in journals spanning a diverse range of scholarly fields, including communication, social media, health, race, and linguistics. Since 2022, three years into the COVID-19 pandemic, there has been a sustained increase in the volume of research on this topic. The studies employed a well-balanced mix of quantitative, qualitative, and mixed-method approaches, with statistical insights and real-life case studies complementing each other. The identified strategies ranged from general social media use to targeted approaches directed at oneself, in-group members, and out-group members, which offer a valuable reference for future research aimed at further expanding and categorizing anti-racism strategies. The studies also showed a balanced scholarly focus on different agents involved in anti-racism efforts for the Asian community.

To advance the study of effective social media-mediated anti-racism strategies for the Asian community, this SLR illuminates the following research directions.

First, more future studies could explore relevant strategies from a broader range of perspectives, such as race, health, and linguistics, extending beyond the predominant focus on communication within an Asian context or on media and communication studies.

Second, this body of scholarship would benefit from more nuanced understandings of

(a) localized anti-racism efforts mediated by social media in non-U.S. contexts, (b) the specific strategies enabled by distinct features of various social media platforms, and (c) anti-racism strategies across a broader range of social and cultural contexts.

Third, further research employing diverse methodological approaches is needed to investigate and triangulate the interrelationships between anti-racism strategies, their agent, effectiveness, and outcomes. Identifying consistent patterns can provide valuable insights for key stakeholders, including policymakers and activist organizations, to implement effective social media-mediated anti-racism strategies. Future studies should also further relevant examination in an expanded pool of strategies, considering the continuously evolving practices of social media users.

Fourth, the challenges associated with using social media to combat racism against the Asian community warrant more systematic examination. These issues extend beyond anti-racism efforts and may significantly limit the effectiveness of social media-mediated initiatives in addressing social justice concerns.

However broad these suggestions may seem, their ultimate purpose is to inspire detailed research agendas focused on a central goal: identifying effective social media-mediated anti-racism strategies for the Asian community. Research in this area ultimately addresses, singly or in combination, three core elements: (1) the linguistic and rhetorical strategies employed, (2) their application through the specific affordances of various social media platforms, and (3) the societal impact they generate. All of these elements are central to the overarching mission of combating racism against the Asian community.

Taking the strategy of assertive coping as an example, future research can explore the linguistic framing users employ when exposing racism-related incidents. For instance, a specific analysis of humorous rhetorical devices, such as satire, parody, and counter-humor, could illuminate their unique persuasive and cathartic functions.

Further, it is critical to investigate how these messages are tailored to the multimodal affordances of various social media platforms. This includes examining, for example, whether the assertive narratives are constructed through short-form videos, text-based posts, memes, or hybrid formats, and whether they are disseminated via original posts or interactions in comment sections.

From a media audience and effects perspective, scholars could examine the impact of these strategies on public discourse. Key questions include: How do they shape representations of the Asian community? To what extent do the assertions spark productive dialogue or foster solidarity across different communities? What is their role in the broader ecosystem of global digital activism?

Pursuing these agendas through interdisciplinary, context-sensitive, and methodologically diverse approaches, as suggested above, will yield rich, nuanced insights into this critical area of scholarship.

Theoretical and empirical implications

Theoretically, this SLR maps the scholarly landscape of how social media's role in advancing social justice has been conceptualized and operationalized in studies of anti-racism strategies. The strategies identified align with the framework proposed by Xin Zhao et al. (2024), which posits that individuals' participation in social justice initiatives via social media

progresses through stages: from awareness, to the collaborative construction and expansion of knowledge, to the critique of underlying structural problems, and finally to mobilizing calls to action. Strategies identified in this SLR echoes the above framework. For instance, consumptive social media use primarily facilitates awareness raising. Expressive use and the exchange of lived experiences aid in collective knowledge construction and expansion. Finally, social media activism is critical for facilitating structural critique and mobilizing calls to action. Furthermore, practices such as self-presenting racial/ethnic identity and assertive coping (toward aggressors) highlight the unique particularities of anti-racism initiatives and their specific leverage of social media affordances to achieve strategic purposes.

This SLR also provides empirical insights to aid policymakers, social media platform designers, and activist organizations in designing targeted and efficient social media-driven anti-racism initiatives for both Asian and non-Asian communities.

First, to improve the psychological well-being of individuals involved in anti-racism efforts, stakeholders could encourage expressive social media use, such as publishing, sharing, or commenting on content concerning anti-racism for the Asian community on social media. This strategy can enhance feelings of empowerment, increase perceptions of social support, and reduce worry about discrimination within the Asian community. Concurrently, it can help reduce negative perceptions of the Asian community among non-members.

Second, to mobilize the Asian community towards concrete change, initiatives can focus on encouraging community members to engage in social media activism that targets structural reform. For instance, members can be supported in organizing and calling for online collective actions, such as petitions, boycotts, or legislative campaigns, to address the root causes of racism. This strategic online engagement is crucial for fostering positive attitudes toward wider political participation and bridging the gap between online advocacy and offline action.

Third, to build and strengthen relational networks, both within the Asian community and with external groups, relevant parties can strategically facilitate the exchange of lived or vicarious experiences with racism. This strategy amplifies the reach and visibility of anti-Asian racism, which can effectively raise public awareness and foster both in-group and cross-community solidarity.

Conclusion

This paper systematically reviewed 38 studies that examined social media-mediated anti-racism strategies for the Asian community. Theoretically, it contributes to the growing scholarly discourse on the role of social media in addressing social justice issues. Empirically, the findings provide insights into the development of effective strategies to empower the Asian community and society at large in combating racism through social media platforms. The relatively small number of studies and their dispersed scholarly focus limited this SLR's ability to draw insights into the applied grounding theories, as is often possible in a more mature field. This limitation, however, underscores the need for further research and highlights the significant potential for growth within this area of study.

References

Abidin, C., & Zeng, J. (2020). Feeling Asian together: Coping with #COVIDRacism on Subtle Asian Traits. *Social Media + Society*, *6*(3), 1–5. https://doi.org/10.1177/2056305120948223

Ahmed, S., Chen, V. H. H., Jaidka, K., Hooi, R., & Chib, A. (2021). Social media use and anti-immigrant attitudes: Evidence from a survey and automated linguistic analysis of Facebook posts. *Asian Journal of Communication*, *31*(4), 276–298. https://doi.org/10.1080/01292986.2021.1929358

Atkin, A. L., Ahn, L. H., Yi, J., & Li, J. (2024). A qualitative study of Asian American adolescents' experiences of support during the COVID-19 and racism syndemic. *Asian American Journal of Psychology*, *15*(4), 295–307. https://doi.org/10.1037/aap0000342

Brondolo, E., Brady, N., Pencille, M., Beatty, D., & Contrada, R. J. (2009). Coping with racism: A selective review of the literature and a theoretical and methodological critique. *Journal of Behavioral Medicine*, *32*(1), 64–88. https://doi.org/10.1007/s10865-008-9193-0

Cai, M., Ahmed, S., Ibasco, G. C., & Chib, A. (2024). Two sides of a coin: Understanding social media use and its relationships to online perceived discrimination and life satisfaction. *Asian Journal of Communication*, *34*(6), 599–617. https://doi.org/10.1080/01292986.2024.2398477

Cao, J., Lee, C., Sun, W., & De Gagne, J. C. (2022). The #StopAsianHate movement on Twitter: A qualitative descriptive study. *International Journal of Environmental Research and Public Health*, *19*(7), 1–11. https://doi.org/10.3390/ijerph19073757

Chaif, R. H., & Finneman, T. (2024). "#My place isn't in the kitchen": Examining feminist Facebook framing of an Algerian social movement. *Social Media + Society*, *July-September*, 1–11. https://doi.org/10.1177/20563051241274657

Cheah, C. S. L., Wang, C., Ren, H., Zong, X., Cho, H. S., & Xue, X. (2020). COVID-19 racism and mental health in Chinese American families. *Pediatrics*, *146*(5), 1–12. https://doi.org/10.1542/peds.2020-021816

Chen, Z. F., Sun, R., & Tao, W. (2024). Channeling engagement into action: The role of empowerment in Asian Americans' social media use in combating anti-Asian discrimination. *Asian Journal of Communication*, *34*(2), 236–257. https://doi.org/10.1080/01292986.2024.2315587

Cho, H., Li, W., Cannon, J., Lopez, R., & Song, C. (2021). Testing three explanations for stigmatization of people of Asian descent during COVID-19: Maladaptive coping, biased media use, or racial prejudice? *Ethnicity & Health*, *26*(1), 94–109. https://doi.org/10.1080/13557858.2020.1830035

Chon, M.-G. (2023). The role of social media in empowering activism: Testing the integrative model of activism to anti-Asian hate crimes. *Asian Journal of Communication*, *33*(6), 511–528. https://doi.org/10.1080/01292986.2023.2251131

Chon, M.-G., & Park, H. (2020). Social media activism in the digital age: Testing an integrative model of activism on contentious issues. *Journalism & Mass Communication Quarterly*, *97*(1), 72–97. https://doi.org/10.1177/1077699019835896

Chou, R. S., & Feagin, J. R. (2016). *Myth of the model minority: Asian Americans facing racism*. Routledge.

Clark, L. S. (2016). Participants on the margins: #BlackLivesMatter and the role that shared artifacts of engagement played among minoritized political newcomers on Snapchat,

Facebook, and Twitter. *International Journal of Communication*, *10*, 235–253.

Corneau, S., & Stergiopoulos, V. (2012). More than being against it: Anti-racism and anti-oppression in mental health services. *Transcultural Psychiatry*, *49*(2), 261–282. https://doi.org/10.1177/1363461512441594

Criss, S., Nguyen, T. T., Michaels, E. K., Gee, G. C., Kiang, M. V., Nguyen, Q. C., … Kennedy, C. J. (2023). Solidarity and strife after the Atlanta spa shootings: A mixed methods study characterizing Twitter discussions by qualitative analysis and machine learning. *Frontiers in Public Health*, *11*, 1–11. https://doi.org/10.3389/fpubh.2023.952069

Edrington, C. L. (2022). Social movements and identification: An examination of how Black Lives Matter and March for Our Lives use identification strategies on Twitter to build relationships. *Journalism & Mass Communication Quarterly*, *99*(3), 643–659.

Ellefsen, R., Banafsheh, A., & Sandberg, S. (2022). Resisting racism in everyday life: From ignoring to confrontation and protest. *Ethnic and Racial Studies*, *45*(16), 435–457. https://doi.org/10.1080/01419870.2022.2094716

Freelon, D., McIlwain, C., & Clark, M. (2018). Quantifying the power and consequences of social media protest. *New Media & Society*, *20*(3), 990–1011. https://doi.org/10.1177/1461444816676646

Hanasono, L. K., & Yang, F. (2016). Computer-mediated coping: Exploring the quality of supportive communication in an online discussion forum for individuals who are coping with racial discrimination. *Communication Quarterly*, *64*(4), 369–389. https://doi.org/10.1080/01463373.2015.1103292

Hannouch, B., & Milstein, T. (2025). Activating ecocentrism: How young women environmental activists produce identity on Instagram. *Environmental Communication*, *19*(2), 198–217. https://doi.org/10.1080/17524032.2024.2376697

Huang, J. T., Krupenkin, M., Rothschild, D., & Cunningham, J. L. (2023). The cost of anti-Asian racism during the COVID-19 pandemic. *Nature Human Behaviour*, *7*, 682–695. https://doi.org/10.1038/s41562-022-01493-6

Jun, J., Kim, J. K., & Woo, B. (2024). Fight the virus and fight the bias: Asian Americans' engagement in activism to combat Anti-Asian COVID-19 racism. *Race and Justice*, *14*(2), 233–250. https://doi.org/10.1177/21533687211054165

Kang, S. (2023). Civic engagement in anti-Asian violence activism: A comparative view between Asians and non-Asian ethnic groups in the United States. *Asian Journal of Communication*, *33*(2), 182–208. https://doi.org/10.1080/01292986.2023.2180528

Keum, B. T., & Volpe, V. (2023). Resisting and countering online racial hate: Antiracism advocacy and coping online with racism as moderators of distress associated with online racism. *Journal of Counseling Psychology*, *70*(5), 498–509. https://doi.org/10.1037/cou0000674

King-O'Riain, R. C. (2022). #Wasian check: Remixing 'Asian + White' multiraciality on TikTok. *Genealogy*, *6*(2), 1–21. https://doi.org/10.3390/genealogy6020055

Kuo, R., & Jackson, S. J. (2024). The political uses of memory: Instagram and Black-Asian solidarities. *Media, Culture & Society*, *46*(1), 164–186. https://doi.org/10.1177/01634437231185963

Lee, D. L., & Ahn, S. (2011). Racial discrimination and Asian mental health: A meta-analysis. *The Counseling Psychologist*, *39*(3), 463–489. https://doi.org/10.1177/0011000010381791

Lee, E. (2016). *The making of Asian America: A history*. New York, NY: Simon and Schuster.

Lee, J. J., & Lee, J. (2023). #StopAsianHate on TikTok: Asian/American women's space-making for spearheading counter-narratives and forming an ad hoc Asian community. *Social Media + Society*, *January-March*, 1–11. https://doi.org/10.1177/20563051231157598

Lee, S., & Waters, S. F. (2021). Asians and Asian Americans' experiences of racial discrimination during the COVID-19 pandemic: Impacts on health outcomes and the buffering role of social support. *Stigma and Health*, *6*(1), 70–78. https://doi.org/10.1037/sah0000275

Lee, Y.-J., Haley, E., & Shang, Y. (2024). Exploring anti-Asian racism activism on Twitter during the early era of COVID-19 hate crimes: Implications for marketers' social purpose communication strategy. *Journal of Current Issues & Research in Advertising*, *45*(1), 88–111. https://doi.org/10.1080/10641734.2023.2252025

Lee-Won, R. J., Lee, J. Y., Song, H., & Borghetti, L. (2017). "To the bottle I go . . . to drain my strain": Effects of microblogged racist messages on target group members' intention to drink alcohol. *Communication Research*, *44*(3), 388–415. https://doi.org/10.1177/0093650215607595

Lough, K., & McIntyre, K. (2023). A systematic review of constructive and solutions journalism research. *Journalism*, *24*(5), 1069–1088.

Manohar, U., & Kline, S. L. (2024). The role of social support in disarming the effects of racial microaggressions. *Communication Research*, *51*(5), 580–603. https://doi.org/10.1177/00936502231151740

Melchior, C., & Oliveira, M. (2022). Health-related fake news on social media platforms: A systematic literature review. *New Media & Society*, *24*(6), 1500–1522. https://doi.org/10.1177/14614448211038762

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097.

Mundt, M., Ross, K., & Burnett, C. M. (2018). Scaling social movements through social media: The case of Black Lives Matter. *Social Media + Society*, *October-December*, 1–14. https://doi.org/10.1177/2056305118807911

Nartey, M. (2022). Centering marginalized voices: A discourse analytic study of the Black Lives Matter movement on Twitter. *Critical Discourse Studies*, *19*(5), 523–538. https://doi.org/10.1080/17405904.2021.1999284

Odağ, Ö., & Moskovits, J. (2024). We are not a virus: Repercussions of anti-Asian online hate during the COVID-19 pandemic on identity and coping strategies of Asian-heritage individuals. *Ethnic and Racial Studies*, 1–32. https://doi.org/10.1080/01419870.2024.2362459

Pan, S., Yang, C., Tsai, J.-Y., & Dong, C. (2021). Experience of and worry about discrimination, social media use, and depression among Asians in the United States during the COVID-19 pandemic: Cross-sectional survey study. *Journal of Medical Internet Research*, *23*(9), 1–10. https://doi.org/10.2196/29024

Park, M., Woo, B., Jung, H.-M., Jeong, E., Choi, Y., Takeuchi, D., & Peregrina, H. N. (2024). COVID-19, racial discrimination and civic engagement among Filipino American and Korean American young adults. *Emerging Adulthood*, *12*(2), 236–251. https://doi.org/10.1177/21676968231224098

Parker, D., & Song, M. (2006). New ethnicities online: Reflexive racialisation and the Internet. *The Sociological Review*, *54*(3), 575–594. https://doi.org/10.1111/j.1467-954X.2006.00630.x

Rivas-Drake, D., Pinetta, B. J., Juang, L. P., & Agi, A. (2022). Ethnic-racial identity as a source of resilience and resistance in the context of racism and xenophobia. *Review of General Psychology*, *26*(3), 317–326. https://doi.org/10.1177/10892680211056318

Shi, L., Zhang, D., Martin, E., Chen, Z., Li, H., Han, X., … Su, D. (2022). Racial discrimination, mental health and behavioral health during the COVID-19 pandemic: A national survey in the United States. *Journal of General Internal Medicine*, *37*(10), 2496–2504. https://doi.org/10.1007/s11606-022-07540-2

Shin, W., Wang, W. Y., & Song, J. (2023). COVID-racism on social media and its impact on young Asians in Australia. *Asian Journal of Communication*, *33*(3), 228–245. https://doi.org/10.1080/01292986.2023.2189920

Tao, W., Li, J.-Y., Lee, Y., & He, M. (2024). Individual and collective coping with racial discrimination: What drives social media activism among Asian Americans during the COVID-19 outbreak. *New Media & Society*, *26*(6), 3168–3187. https://doi.org/10.1177/14614448221100835

Tong, S. T., Stoycheff, E., & Mitra, R. (2022). Racism and resilience of pandemic proportions: Online harassment of Asian Americans during COVID-19. *Journal of Applied Communication Research*, *50*(6), 595–612. https://doi.org/10.1080/00909882.2022.2141068

Uyheng, J., Bellutta, D., & Carley, K. M. (2022). Bots amplify and redirect hate speech in online discourse about racism during the COVID-19 pandemic. *Social Media + Society*, *8*(3), 205630512211047. https://doi.org/10.1177/20563051221104749

Wang, S., Chen, X., Li, Y., Luu, C., Yan, R., & Madrisotti, F. (2021). 'I'm more afraid of racism than of the virus!': Racism awareness and resistance among Chinese migrants and their descendants in France during the Covid-19 pandemic. *European Societies*, *23*(sup1), S721–S742. https://doi.org/10.1080/14616696.2020.1836384

Wang, X., Wu, M., & Rajtmajer, S. (2023). From Yellow Peril to Model Minority: Asian stereotypes in social media during the COVID-19 pandemic. *Proceedings of the 15th ACM Web Science Conference 2023*, 283–291. Austin TX USA: ACM. https://doi.org/10.1145/3578503.3583614

Willnat, L., Shi, J., & De Coninck, D. (2023). Covid-19 and xenophobia in America: Media exposure, anti-Asian stigmatization, and deservingness of Asian immigrants. *Asian Journal of Communication*, *33*(2), 87–104. https://doi.org/10.1080/01292986.2023.2176898

Yang, C., & Tsai, J.-Y. (2023). Asians and Asian Americans' social media use for coping with discrimination: A mixed-methods study of well-being implications. *Heliyon*, *9*(6), 1–16. https://doi.org/10.1016/j.heliyon.2023.e16842

Yang, F., & Hanasono, L. K. (2021). Coping with racial discrimination with collective power: How does bonding and bridging social capital help online and offline? *Howard Journal of Communications*, *32*(3), 274–293. https://doi.org/10.1080/10646175.2021.1910882

Yu, N., Pan, S., Yang, C., & Tsai, J.-Y. (2020). Exploring the role of media sources on COVID-19–related discrimination experiences and concerns among Asian people in the United States: Cross-sectional survey study. *Journal of Medical Internet Research*, *22*(11), e21684. https://doi.org/10.2196/21684

Zhao, Xin, Feigenbaum, A., & Demirkol Tønnesen, Ö. (2024). Between comments and collective action: The potential of TikTok in endometriosis advocacy. *International Journal of Communication*, *18*, 3611–3633.

Zhao, Xin, Feigenbaum, A., & McDavitt, S. (2022). Feasibility of comics in health communication: Public responses to graphic medicine on Instagram during the COVID-19 pandemic. *Journal of Visual Political Communication*, *9*(1), 9–28. https://doi.org/10.1386/jvpc_00015_1

Zhao, Xin, & Xiang, Y. (2023). Using disparagement humour to deal with health misinformation endorsers: A case study of China's shuanghuanglian oral liquid incident. In K. Fowler-Watt & J. McDougall (Eds.), *The Palgrave Handbook of Media Misinformation* (pp. 179–190). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-11976-7_12

Zhao, Xinyu, & Abidin, C. (2023). The "Fox Eye" challenge trend: Anti-racism work, platform affordances, and the vernacular of gesticular activism on TikTok. *Social Media + Society*, *January-March*, 1–16. https://doi.org/10.1177/20563051231157590

Zhu, H. (2020). Countering COVID-19-related anti-Chinese racism with translanguaged swearing on social media. *Multilingua*, *39*(5), 607–616. https://doi.org/10.1515/multi-2020-0093

Xin Zhao is a Principal Lecturer in Media and Communication at Bournemouth University, UK. Her research focuses on the role of media representations, media audiences, and journalistic practices in constructing and/or challenging social justice issues. Her recent work is particularly dedicated to exploring how media and communication can address racism against the Asian community. She has published in top journals including Journalism, Journalism Studies, Journalism Practice, Digital Journalism, Health Communication, International Journal of Communication, Asian Journal of Communication, and Global Media and Communication. Her research has been supported by the British Academy.

This work was supported by the British Academy/Leverhulme Small Research Grants [grant number SRG24\240850].

Journal of Language, Media and Society
Vol. 1, No. 1 https://doi.org/10.66056/jlms.2026.v1i1.002

Carter & Co. PUBLISHING

WHEN SYSTEMS MISRECOGNIZE THEIR USERS: A SEMI-SYSTEMATIC REVIEW OF COMMUNICATION, IDENTITY, AND BIAS IN LLMS

LIJING GAO & RUANJIA LIU

Large language models (LLMs) offer vast computational power yet consistently overlook cultural nuance. While often praised for bridging language gaps, recent research highlights a deeper issue: LLMs largely reflect Anglophone and Western European cultural values, which are embedded in the English-dominated data that shapes them. This review synthesizes findings from 2020 to 2025 to assess the implications of this technological spread for vulnerable groups, particularly immigrants, refugees, and international students, who must navigate adaptation within English-centric and Western communication norms.

Using insights from cultural cognition and identity-protective cognition theories, the analysis identifies five central forms of bias: (1) representational bias that undermines non-Western perspectives; (2) linguistic inequity that amplifies challenges for low-resource languages; (3) authenticity failures, with stereotypes substituting for real cultural understanding; (4) identity erosion as users' voices are homogenized; and (5) reliance on LLMs that may hinder independent language skill development. This "equity paradox" means that the very systems marketed as democratizing global communication can actually deepen exclusion and sameness among those who are most reliant on them.

Ultimately, the review concludes that current governance and policy efforts are insufficient to address the underlying power dynamics that shape LLM development. Authentic cross-cultural communication, the evidence suggests, depends on human qualities absent in LLMs: presence, vulnerability, and the openness to change that underpins accurate understanding. In an AI-mediated world, recognizing the limits of these tools is not a matter of nostalgia, but rather necessary wisdom.

KEYWORDS: large language models; cross-cultural communication; cultural bias; identity formation; linguistic diversity; cultural homogenization; digital equity; AI governance; cultural representation

Introduction: The Seduction and Betrayal of Technological Solutions

Current discussions regarding large language models (LLMs) frequently resemble a form of technological idealism, as innovative computer tools pledge to eliminate the obstacles that have historically hindered human interaction across linguistic and cultural divides. The narrative is compelling: LLMs such as ChatGPT and Claude serve serve as democratic instruments, accessible worldwide, enabling nonnative speakers to communicate with unparalleled fluency, tapping into cultural insights and linguistic refinement that were once attainable solely

© 2026 Carter & Co.

through extensive immersive experience (Hu et al., 2025; Khan et al., 2025; Li et al., n.d.; Lo et al., 2024; Mao et al., 2025; Zangana et al., 2025; Zhang et al., 2025; Zohouri et al., 2024). From this perspective, LLMs appear almost redemptive, promising to correct historical inequities in global communication systems.

Yet this narrative conceals more troubling realities. Behind the outer layer of acceptable performance metrics lies an intrinsic bias that privileges certain viewpoints, assessments, and communication styles while marginalizing others (Ashraf et al., 2025; Keleg, 2025; Shan et al., 2024). The algorithms being implemented worldwide were predominantly trained on English-language materials authored mainly by Anglophone writers, reflecting Anglophone concerns. When asked to respond as an "average person," GPT models align most closely with cultural values from Finland, Andorra, and the Netherlands while showing the greatest distance from Jordan, Libya, and Ghana (Dairo, 2005; Ferdaus et al., 2024; Liu et al., 2025; Lyu & Du, 2025). These are not random variations. They represent systematic underrepresentation of non-Western moral frameworks, epistemologies, and ways of understanding the world.

For newcomers navigating life in a foreign country, whether immigrants working to understand local social norms, refugees grappling with cultural adaptation, or international students seeking connection in unfamiliar settings, these technologies present a particularly intricate set of challenges. On one side, they provide authentic support: instruments that aid non-native speakers in generating more fluent English, alleviating the cognitive load of translation, and expediting language acquisition (Ehrensberger-Dow et al., 2020; Elgamal, 2019). Conversely, these instruments may consistently misrepresent local culture, perpetuate detrimental preconceptions, standardize unique expressions, and foster dependencies that compromise the independent proficiency necessary for enduring cultural adaptation (Algouzi & Alzubi, 2023; Sahebi & Formosa, 2025).

This article presents a semi-systematic review of the dual promise and peril of large language models in cross-cultural contexts. The review synthesizes emerging evidence showing that cultural bias is not a peripheral flaw but an intrinsic feature of these systems to shaped by their training data, parameter settings, optimization processes, and institutional uses. Across the literature, a consistent pattern emerges regarding who benefits from current deployment practices and who disproportionately bears the costs. At its core, this review asks what genuine equity would require in AI-mediated communication.

Semi-Systematic Review Rationale and Methods

Rationale

This review employs a semi-systematic methodology, which combines the transparency and reproducibility of systematic review protocols with the methodological flexibility necessary for research in rapidly evolving domains (Snyder, 2019). Unlike fully systematic reviews that require exhaustive search strategies and rigid adherence to all PRISMA guidelines (Page et al., 2021), semi-systematic approaches allow targeted inclusion criteria and pragmatic methodological adjustments, particularly well-suited to interdisciplinary and fast-moving fields such as large language model research (Snyder, 2019). This approach balances the need for evidence synthesis rigor with recognition that emerging research areas may benefit from focused rather than exhaustive searches.

Inclusion Criteria

Studies were eligible for inclusion referred to the following pre-specified criteria, established following PRISMA 2020 guidelines (Page et al., 2021):

Publication timeframe: Studies published between January 2020 and October 2025 were included. This five-year window captures the post-ChatGPT era beginning with its November 30, 2022 release (OpenAI, 2022), which catalyzed unprecedented research interest in generative artificial intelligence. All literature searches were conducted across multiple databases, including Google Scholar, Web of Science, and arXiv. While Web of Science and arXiv returned relevant results, Google Scholar provided the most comprehensive coverage by aggregating papers from diverse sources relevant to this research. A five-year search window captures rapidly evolving fields through major technological transitions while remaining operationally feasible. (Hamman et al, 2017).

Study design: Empirical research was prioritized, including studies examining LLM use in cross-cultural or multilingual contexts as well as broader research on cross-cultural communication and cultural adaptation challenges that provide contextual framework for understanding LLM applications. Preprints and conference proceedings were included to ensure a comprehensive and timely evidence synthesis in this rapidly evolving field. In fast-moving technological domains, traditional peer-review timelines can render findings outdated before publication (Suh, 2025). The COVID-19 pandemic established clear precedent for incorporating preprints into systematic reviews, showing that preprints exhibit minimal discrepancies from their later published versions and can improve estimate precision while enabling timely dissemination (Tennant et al., 2018). In AI research in particular, platforms such as arXiv now serve as central communication channels (Suh et al., 2025), and major institutions and technology companies routinely use preprints to accelerate knowledge transfer (Suh et al., 2025; Sætra, 2024). This practice aligns with open science principles that prioritize transparency, accessibility, and rapid dissemination (Tong et al., 2025; UNESCO, 2021), which are particularly critical for evidence synthesis addressing contemporary AI applications where policy and practice decisions require current information (Yang et al., 2025). Conference proceedings were included because they provide access to emerging research and diverse perspectives that may not yet appear in journal publications, helping reduce publication bias and identify relevant evidence comprehensively (Scherer & Saldanha, 2019). Empirical evidence to defined as research based on observed and measured phenomena rather than theory or belief (Cook et al., 1997) to provides the strongest foundation for evidence synthesis and supports valid conclusions regarding real-world applications (Gopalakrishnan & Ganeshkumar, 2013; Tong et al., 2025). This inclusive approach allows integration of relevant contextual research that clarifies the cultural and communicative dimensions within which LLMs operate, while maintaining methodological rigor through systematic screening and quality assessment procedures.

Language: English-language studies were included due to resource constraints. This pragmatic inclusion criterion reflects the semi-systematic approach, which balances rigor with operational feasibility (Snyder, 2019; Pham et al., 2005), though the resulting language bias is acknowledged as a study limitation.

Exclusion Criteria

Studies were systematically excluded if they: (1) lacked cultural analysis (purely technical or algorithmic papers without examination of cultural, linguistic, or social dimensions); (2) were non-empirical (opinion pieces, commentaries, editorials, or theoretical work without original data); (3) examined exclusively monolingual contexts as the review focuses specifically on cross-cultural and multilingual dimensions of LLM use and cultural adaptation unless they provided empirical evidence on communication or cultural adaptation challenges directly applicable to understanding LLM performance in cross-cultural contexts. As a result, this research identified 536 articles into analysis.

The methodological workflow for this semi-systematic review is depicted in Figure 1, highlighting the pragmatic integration of PRISMA guidelines with targeted inclusion strategies to synthesize empirical research from January 2020 to October 2025.

Figure 1: Semi-Systematic Review Methodology and Study Selection Flow

Phase	Action & Criteria Details
1. Identification	**Database Search (Jan 2020 – Oct 2025)** Sources: Google Scholar (Primary), Web of Science, arXiv. Focus: Post-ChatGPT era & technological transitions.
2. Screening	**Inclusion Criteria (Pragmatic & Targeted)** **Type:** Empirical research, preprints, & conference proceedings. **Scope:** LLMs in cross-cultural/multilingual contexts; cultural adaptation. **Language:** English only (operational feasibility).
3. Eligibility	**Exclusion Criteria** 1. **Technical Only:** Lacks cultural/social analysis. 2. **Non-Empirical:** Editorials, opinions, or purely theoretical work. 3. **Monolingual:** Excludes studies without cross-cultural dimensions.
4. Included	**Final Sample Size: N = 536 Articles** Synthesized to address three core Research Questions (RQs).

This semi-systematic review examines how large language models mediate cross-cultural communication and in which contexts their limitations become most consequential. Drawing on the evidence gathered, the review is structured around three central research questions: **(1)** How do LLMs facilitate or hinder cross-cultural and multilingual communication for users navigating unfamiliar cultural contexts? **(2)** What cultural biases and worldview misalignments are systematically embedded in LLM outputs, and how do these patterns shape users' trust,

interpretation, and reliance on these systems? **(3)** What would genuine communicative equity require in the design, training, and governance of LLMs used across diverse cultural settings? These questions organize the synthesis that follows and provide a framework for interpreting patterns across the empirical literature.

Results

Initial analysis of the corpus confirms that the functional benefits of LLMs for cross-cultural communication are substantial and widely documented. For users operating outside their dominant language, these systems act as critical cognitive scaffolds, significantly reducing the mental burden of translation and syntactic formulation (Lee et al., 2024). By elevating lexical precision and smoothing linguistic roughness, LLMs effectively democratize access to high-stakes professional and academic discourse, offering a provisional form of linguistic equity that allows newcomers to bypass traditional gatekeeping mechanisms (Nguyen et al., 2024; Liang et al., 2023). However, the literature increasingly suggests that this improved surface-level accessibility frequently comes at the cost of deeper semantic integrity.

Section I: The Architecture of Bias—How Cultural Representation Becomes Cultural Erasure

1.1 The Manifestation of Western-Centrism

One might begin with a simple question: What does a language model "believe"? The answer reveals itself not in explicit statements but in consistent patterns of preference and omission. Systematic evaluation across diverse LLM architectures reveals a highly consistent pattern across the majority of studies: when asked to engage with cultural questions, these systems tend to prioritize values characteristic of Protestant European and Anglo-American contexts (Segerer, 2025).They default to self-expression values such as environmental protection, tolerance of diversity, and gender equality, which predominate in wealthy Western societies, even when prompted in non-Western languages or explicitly instructed to adopt non-Western perspectives. In what follows, we use a series of illustrative, composite scenarios to concretize patterns observed across the empirical literature. These examples are not single case reports but theoretically and empirically informed vignettes that synthesize recurring dynamics identified in prior studies.

This is not neutral linguistic performance. It is cultural transmission. When LLMs repeatedly associate Indian women with domestic labor, Russians with vodka consumption, or Arabs with terrorism, they do not merely reflect societal biases that happened to appear in training data (Wu et al., 2025). Rather, they crystallize and scale those biases across millions of interactions, routinizing and legitimizing them through the appearance of technological objectivity.

Consider the consequences for newcomers. Unfamiliar with local social conventions such as professional clothing, conversational indirectness, and friendship duties, an immigrant confronts an LLM schooled mostly in Western models and frameworks. When she asks about dinner etiquette or gift-giving customs, she receives advice filtered through Western preoccupations and Western categorical systems (Pedersen et al., 2025). The system has little understanding of her destination culture because few literature in that language describe it

well enough to modify model parameters. Instead of silence or doubt, she receives information that appears confidence, cultural expertise, and neutral information. Only after social failures and misunderstandings does she realize the advice was misleading.

The mechanisms that contribute to the social cohesion bias are well understood. Large language models are trained on textual data, which is significantly unbalanced within global information systems. English prevails, despite accounting for approximately 15% of the global population (Crystal, 2003; Rao, 2019). Western perspectives prevail, even though they account for a diminutive fraction of human diversity. The training data reflects this imbalance, not due to intentional decisions but as a result of the inherent structures of digital information distribution. Once established, this imbalance proves to be notably challenging to remove. Models retain these biases, even when trained on multilingual data (Ashraf et al., 2025; Keleg, 2025). Despite explicit instructions to incorporate non-Western perspectives, they process requests through frameworks influenced by their training. The bias is an inherent characteristic of the architecture rather than a correctable error.

1.2 The Hidden Mechanisms: How Bias Persists Despite Efforts to Address It

Understanding cultural bias in large language models requires examining mechanisms operating at multiple levels, often invisible to users and insufficiently addressed by current interventions. This review traces the process through which bias is generated and perpetuated to from imbalances in training corpora and tokenization, through model objectives and alignment procedures, to interface design and institutional deployment practices that normalize some voices while marginalizing others.

Data Composition and Representational Asymmetry

Across the studies reviewed, there is broad and consistent evidence that the foundational source of cultural bias in large language models stems from a profound asymmetry in their training corpora. Multiple analyses show that these datasets are dominated by English-language texts drawn from Western media, publishing, and digital platforms (Ferdaus et al., 2024; Ghimire, 2025; Han et al., 2025; Liu et al., 2025; Lyu & Du, 2025). The imbalance is striking: although English accounts for roughly 15 percent of global languages, it makes up an estimated 50 to 70 percent of the digitized text used to train major LLMs (Segerer, 2025; Lehdonvirta, 2022; Chatterji et al. 2025). In contrast, low-resource languages, typically those with fewer than ten million speakers and limited digital documentation, contribute less than five percent of most training datasets, even though they represent the home languages of billions of people (Ferdaus et al., 2024; Liu et al., 2025; Lyu & Du, 2025).

The problem extends beyond quantitative imbalance to what might be described as a representational hierarchy. The texts included in training corpora reflect not only the distribution of languages but also the knowledge systems that are recognized as authoritative. Academic publishing, news media, and digitized books, which serve as the primary sources of training data, are concentrated in Western institutions and largely encode Western epistemological frameworks. Indigenous knowledge systems, non-Western philosophical traditions, and alternative ways of understanding fundamental concepts such as causality, time, personhood, and ethics are either absent from training data or appear only when mediated through Western interpretive frameworks and scholarly representations (Abdilla & Crawford,

2020; Birhane et al., 2022; De Sousa Santos, 2014; Gwagwa et al., 2020; Hoppers, 2002; Kamran, 2024; Leibo et al., 2025; Peters & Carman, 2024). For example, a Navajo environmental practice may be recorded not by Navajo knowledge holders but by Western environmental scientists describing it; the model therefore learns the Western account rather than the Indigenous understanding. This compositional bias means that the model is trained not on the full diversity of global human knowledge but on a particular, narrowly Western subset that is presented as universal.

When large language models are trained on these compositionally biased datasets, the statistical regularities within the data become encoded not as explicit rules or retrievable statements but as learned parameters distributed across millions of artificial neurons. This marks a critical point of translation: the cultural preferences embedded in the data are mathematically instantiated in the model's functional architecture. Consider concretely what this means. When training data disproportionately associates certain concepts with certain cultural contexts, for example linking leadership with masculine pronouns and Western business terminology (Garg et al., 2018; Müller et al., 2025), associating subsistence practices with non-Western peoples (Malu, 2025), or repeatedly pairing development with Western-style industrialization (Malu, 2025), these associations accumulate into learned representations. The model develops what might be called vector space preferences: patterns in how concepts relate to one another in high-dimensional geometric space (Schröder et al., 2024; Müller et al., 2025). Large language models are 3-6 times more likely to recommend occupations that stereotypically align with a person's perceived gender, with boys receiving substantially more STEM-related career suggestions than girls (Torres et al., 2023; Fock & Siller, 2025). These patterns become part of how the model understands the world, functioning not as programmed bias but as learned associations that disadvantage marginalized groups through statistical co-occurrence patterns in training corpora (Bender et al., 2025; Torres et al., 2023).

The particular challenge with this encoding is that these learned parameters are not easily inspected, modified, or removed. A programmer cannot simply edit individual parameters the way they might edit explicit rules in traditional software. The biases are emergent properties of millions of parameters working together, making them what scholars call implicit knowledge to knowledge that shapes behavior but isn't easily articulated or adjusted. When researchers attempt to debias models after training, they face the problem that the bias isn't localized to one parameter or one layer but distributed throughout the network. Attempts to reduce specific stereotypes through post-training techniques often fail because the underlying statistical patterns remain embedded in the model's fundamental structure (Glickman & Sharot, 2025). To meaningfully address bias at the parameter level often requires architectural redesign or retraining, which are computationally expensive and practically difficult to implement at scale.

Prompt conditioning activates and adjusts these priors without altering their essential nature. Users trying to direct models towards more culturally sensitive outputs through rigorous prompting face a limitation: the model can modify emphasis and nuance, but cannot beyond the inherent viewpoints ingrained in its training (Agarwal et al. 2025; Liang et al; 2025; Shen et al., 2025). An Arabic speaker prompting in Arabic continues to receive outputs that exhibit Western biases, as the model's depiction of Arabic culture has been influenced by English-language texts regarding Arab culture, which are filtered through Western frameworks and frequently perpetuate Orientalist stereotypes (Alyafeai et al., 2023; Sallam & Mousa, 2024).

Human feedback loops often amplify the problem rather than resolve it. When

companies apply reinforcement learning from human feedback, a process in which human evaluators rate model outputs to guide further training, a critical question emerges: who are these humans? The research is clear: they are disproportionately Western, educated, English-speaking (Lodoen & Orchard, 2025). Their preferences become embedded in the model's objectives. The system learns not to produce correct or culturally appropriate responses, but to produce responses that align with the preferences of these specific human evaluators. Bias becomes recursively reinforced (Wang et al., 2024; Glickman & Sharot, 2025).

Downstream adoption institutionalizes and scales the bias. When universities, employers, and government agencies deploy these systems, they do not simply use a tool to they enact and perpetuate the model's embedded cultural assumptions (Bao et al. 2025; Prakash et al. 2025; Zheng, 2024; Garcia, 2025). Students learn using systems that treat Western ways of thinking as standard and other ways as exotic. Employees write using tools that systematically alter their distinctive voices toward Western norms. Citizens seek government information from systems trained on Western legal and administrative frameworks.

The result is an escalating process wherein cultural bias, far from being a marginal concern, sits at the very heart of how these systems operate. Addressing it requires not tinkering at margins but fundamental rethinking of how models are trained, evaluated, and deployed.

Section II: The Lived Experience—What Bias Means for Newcomers

2.1 Cultural Misrepresentation and the Loss of Meaning

A growing body of research demonstrates that LLMs routinely reshape nonnative speakers' writing in ways that privilege Western academic styles and epistemic priorities. The scenario of an Iranian graduate student refining her explanation through ChatGPT reflects a pattern observed across multiple empirical evaluations: while the language becomes more fluent, the system often redirects emphasis, reframes arguments, or omits culturally grounded reasoning. The resulting text is polished yet noticeably aligned with Western communicative conventions, illustrating a phenomenon repeatedly documented in cross-cultural studies of LLM use.

This entails cultural misrepresentation, albeit through subtle, nearly imperceptible modifications. Pedersen and colleagues' research demonstrated that LLMs consistently struggle with the interpretation of culture-specific metaphors and idioms (Pedersen et al., 2025). When prompted to elucidate a Danish phrase grounded in Danish history and culture, both ChatGPT and Llama encountered difficulties, either wrongly incorporating English metaphors or resorting to ambiguous generalizations. The sentiment becomes "lost in translation," not due to linguistic disparities but because the model lacks the historical and cultural acumen required to comprehend how that word encapsulates an entire worldview.

For newcomers, the consequences are compounded. An immigrant trying to understand what it means to "work hard" in a new culture does not encounter the nuanced, context-specific meanings that exist in reality, but instead a simplified, Westernized narrative, which centered on the Horatio Alger story and the mythology of pulling oneself up by the bootstraps. This narrative contrasts sharply with how work and effort are understood in many non-Western cultures, where collective responsibility, family honor, or spiritual purpose may take precedence. Yet the large language model presents its version as a neutral and factual account.

The failure extends beyond metaphor. Multilingual proficiency does not ensure cultural representation: research shows AI systems marginalize African cultural expressions through Eurocentric training data (Bignotti, 2025), Latin American media lack localized AI imagery despite multilingual coverage (Sanguinetti & Palomo, 2025), and meaningful representation requires community-grounded evaluation rather than technical capability alone (Qadri et al., 2025). U.S.-centric bias persisted even when models were prompted in their native languages. The researchers demonstrated that self-consistency was a stronger predictor of intercultural alignment than multilingual competence alone, suggesting that the problem lies not in linguistic translation but in the deeper cultural frameworks guiding how the model processes information.

2.2 Stereotype Reinforcement and the Crystallization of Prejudice

Across the studies included in this review, there is substantial and converging evidence that LLMs reinforce culturally embedded stereotypes in ways that meaningfully influence downstream behavior. Pareek (2025) assessed prominent LLM systems with psychology-based metrics particularly formulated to identify biases. Their findings indicate that even systems intentionally meant to be "value-aligned" displayed systematic preconceptions related to race, gender, religion, health, and other variables.

The troubling part: these word association biases proved diagnostic of downstream discriminatory behavior. Models that showed stereotype bias in controlled tests also generated more biased, inappropriate, or harmful outputs in real-world applications. Stereotypes do not exist harmlessly in latent space. They shape what the system produces when deployed.

For a refugee from Myanmar seeking to understand his place in his new community, these stereotypes are not abstract concerns. When the LLM he consults for advice about workplace relationships consistently associates his background with certain traits or capabilities, whether explicitly or through subtle framing, those biases shape both what information he receives and how he begins to perceive himself. Research on stereotype threat reveals that these impacts are not purely individual; they accumulate through repeated interactions with systems that encode and perpetuate prejudice (Wang et al., 2024; Vasista et al., 2025).

Researchers documented an even more troubling phenomenon: stereotypes can arise spontaneously during LLM-based multi-agent interactions, even when the individual agents begin without any predefined bias. The strength of these stereotypes increases within hierarchical systems and through repeated exchanges (Guo & Xu, 2025; Mehdizadeh & Hilbert, 2025), following the same dimensions of warmth and competence that appear consistently across architectures such as GPT, Claude, Mistral, DeepSeek, and Gemini (Guo & Xu, 2025; Borah & Mihalcea, 2024). These findings suggest that stereotype formation is not a model-specific artifact but a structural feature of how large language models learn from and interact with one another (Haase & Pokutta, 2025; Binkyte, 2025). Addressing these emergent biases requires value alignment frameworks that account for multi-agent dynamics rather than focusing solely on individual model behavior (Zeng et al., 2025).

2.3 Communication Style Mismatch and the Imposition of Inappropriate Norms

Although studies vary in their methodological approaches, a clear pattern emerges

across the literature: LLMs often misalign with the communication norms of cultures that rely on indirectness, relational cues, or high-context signaling. Havaldar and colleagues' Culturally-Aware Conversations framework identifies three dimensions that shape communication: situation, relationship, and cultural background (Havaldar et al., 2025).

Large language models perform well in direct, low-context cultures such as the United States and the Netherlands but struggle in societies where indirectness signals respect, humility outweighs self-promotion, and formality conveys attentiveness. Trained mainly on Western norms, they promote behaviors that conflict with local expectations.

Section III: The Equity Paradox—When Tools for Empowerment Become Mechanisms of Exclusion

3.1 The Cruel Irony: When Non-Native Speakers Must Use AI to Avoid AI Accusations

Recent research revealed a troubling finding: AI detection methods incorrectly label 61.3% of essays authored by non-native English speakers as AI-generated (Liang et al., 2023; Jiang et al., 2024). The mechanism is straightforward: non-native authors inherently exhibit reduced linguistic diversity, diminished syntactic complexity, and decreased lexical richness compared to native speakers to the same qualities characteristic of AI-generated writing (Fraser et al., 2025; Lege, 2024). Detection systems, unable to discriminate between sources of reduced complexity, classify human effort as artificial.

The cruelty of this situation is clear: to avoid false accusations of AI use, non-native writers are compelled to use AI to augment their linguistic diversity. To be recognized as genuinely human, they must first become augmented by AI. This technical paradox threatens the inclusion of non-native English speakers in global academic and professional spheres precisely when access becomes most critical (Jiang et al., 2024; Lege, 2024).

Liang and colleagues quantified this paradox: when ChatGPT improved TOEFL essays to resemble native-speaker writing, the average false positive rate dropped by 49.7%, from 61.3% to 11.6% (Liang et al., 2023). The implication is clear: linguistic diversity functions as a proxy for human authenticity in detection systems, yet non-native speakers inherently lack that diversity due to their ongoing development of linguistic competence (Fraser et al., 2025).

3.2 The Double Jeopardy: Cost Barriers and Performance Degradation

Across diverse technical and empirical studies, there is strong evidence that LLMs systematically disadvantage low-resource languages. Speakers of low-resource languages incur costs that are 4-6 times higher per unit of usage compared to English speakers (Solatorio et al., 2024). This situation illustrates the mechanics of tokenization: languages underrepresented in training data are divided into numerous separate tokens, each consuming additional computational resources (Ahia et al., 2023; Petrov et al., 2023). The result is higher costs paired with diminished returns.

Concurrently, the performance of LLMs significantly diminishes for low-resource languages (Petrov et al., 2023; Rahman et al., 2024). Translation quality metrics reveal substantial performance declines in low-resource languages (Solatorio et al., 2024; Pakray et al., 2025). Healthcare information obtained from LLMs in non-English languages demonstrates

significantly lower accuracy and utility: correctness decreases by 18% in Spanish, Chinese, and Hindi compared to English, with non-English responses 29% less consistent than their English counterparts (Chandra & Jin, 2024). This creates what researchers term double jeopardy to individuals with the fewest resources face the greatest obstacles and poorest outcomes, fundamentally undermining equitable technology access (Solatorio et al., 2024).

3.3 Detection Bias and Response Quality Disparities

Across the literature reviewed, multiple independent studies provide compelling evidence that LLM outputs shift systematically in response to linguistic identity cues and cultural context. A Nature study using matched-guise prompts found that models make systematically more negative decisions for text written in African American English, despite no overt mention of race (Hofmann et al., 2024). A cross-country audit in *PNAS Nexus* showed that default outputs cluster toward English-speaking/Protestant European values, with cultural prompting only partly reducing this bias (Tao et al., 2024). Randomized trials also document anchoring effects in LLMs and show that outputs move when primed by preceding information, underscoring how seemingly minor cues can alter model quality and judgments (Nguyen, 2024). Together, these results indicate that disparities reflect dialect and cultural alignment rather than merely "nativeness," and can be triggered by simple primes.

3.4 The Workplace Dynamics: When Quality Degradation Becomes Social Stigma

Although the published literature on workslop, meaning superficially polished but substantively weak AI-generated workplace content, is still developing, the available evidence shows a consistent and widely recognized pattern of social and organizational consequences. Early studies report that AI-mediated, low-quality responses erode trust among coworkers, requiring recipients to verify accuracy, rewrite vague sections, and navigate the interpersonal discomfort of questioning a colleague's contribution (Hancock & Niederhoffer, 2025). Broader organizational research reinforces these concerns: approximately 40% of workers report rising workloads associated with such breakdowns (Niederhoffer et al., 2025; Richardson & Antonello, 2022), which reliably undermines trust and reduces willingness to collaborate. These burdens fall disproportionately on non-native speakers, who rely on LLMs for linguistic fluency (Brynjolfsson et al., 2025) but whose use of AI tools may be misinterpreted as lack of effort or cultural awareness (Nguyen et al., 2024). Taken together, the emerging evidence indicates that technologies intended to support communication can inadvertently generate stigma and exacerbate inequities (Niederhoffer et al., 2025; Koo, 2025).

3.5 Educational Performance Gaps and Linguistic Brittleness

Across the studies examined, there is **substantial and recurring evidence** that equity concerns emerge in educational settings where students rely on large language models as learning tools. While some models demonstrate relatively stable performance across multiple languages, numerous evaluations report a marked drop in accuracy and instructional quality when the models operate in underrepresented or low-resource languages. Kwak and Pardos (2024), for example, document consistent underperformance for languages such as Irish

and Marathi compared to English-based educational taxonomies, reflecting a broader pattern identified across the literature.

Rodrigues and colleagues evaluated LLM performance in answering educational questions in Brazilian Portuguese across different question types, subjects, and difficulty levels (Rodrigues et al., 2025). Their preliminary findings suggest potential for LLMs to support diverse educational needs, though performance varies by question characteristics. Separate research has demonstrated that LLM performance in English consistently surpasses that of other languages (Solatorio et al., 2024; Ahia et al., 2023), and that employing prompts in English often produces better outcomes even in non-English contexts, indicating that LLM representations are primarily shaped by English-centric training frameworks (Tao et al., 2024).

Recent research reveals significant concerns about LLM brittleness: model performance varies dramatically (±23%) across common benchmarks when only single characters such as delimiters between examples are modified, despite semantic information remaining identical (Su et al., 2025). This brittleness is not unique to minute formatting changes; LLMs show unpredictable performance across prompt phrasing, semantic structure, and element ordering on even elementary tasks like set membership queries (Hergert et al., 2025). While humans exhibit some sensitivity to prompt modifications, LLMs demonstrate greater instability, particularly to typographical errors and label order reversals (Li et al., 2025).

Current benchmarks emphasize standardized, formal writing that inadequately reflects diverse human communication styles, resulting in limited external validity for real-world performance assessment. This brittleness has profound consequences for immigrant populations, international students, and multilingual professionals: AI-mediated guidance and evaluation systems affect academic achievement, job opportunities, and social integration when applied across varied linguistic and stylistic contexts (Su et al., 2025; Hergert et al., 2025).

Section IV: Identity Erasure—How Technology Homogenizes Culture

4.1 The Systematic Stripping of Identity Markers

Language functions simultaneously as a communicative instrument and a primary marker of cultural affiliation and self-perception, profoundly shaping how immigrants negotiate identity in new sociocultural environments (Darginavičienė, 2023; Fielding, 2021; Gao, 2021; Hsiao, 2021). For newcomers, this identity negotiation involves constant balancing: acquiring host-country language and cultural norms while preserving heritage and traditions. This ongoing process shapes psychological well-being, social relationships, and everyday experiences (Joubert & Sibanda, 2022; Kiramba & Oloo, 2023).

However, a broad body of scholarship provides robust evidence that LLM-assisted writing introduces a new complication. When Sourati and colleagues (2025) analyzed LLM text revision, they found that semantic content remains preserved while stylistic elements such as personal voice, cultural markers, and individual language choices undergo systematic alteration toward dominant patterns in training data. When a Senegalese immigrant writes in French with characteristic phrases, cultural references, and linguistic patterns reflecting her heritage, LLM "improvement" strips those markers away, homogenizing her writing toward standardized global norms (Sourati et al., 2025; Kuteeva & Andersson, 2024). Her voice does not merely change, it disappears.

This erasure becomes particularly consequential in academic contexts. While ChatGPT

enhances lexical complexity in non-native English speakers' writing, this linguistic "equalization" paradoxically erases distinctive voice and cultural expression (Lin et al., 2025). The writing appears professionally polished yet experientially inauthentic, and it becomes improved by external metrics while internally fractured. For newcomers already navigating tensions between heritage and integration, LLM-assisted "improvement" intensifies authenticity concerns because the communication is fluent but no longer recognizably their own (Gao, 2021; Kiramba & Oloo, 2023; Ozbek-Damar, 2025). The tools designed to facilitate integration simultaneously eliminate the linguistic markers through which cultural identity is expressed and transmitted (Sourati et al., 2025).

4.2 Linguistic Homogenization and the Erosion of Heritage Language Vitality

Recent scholarship indicates increasing concern regarding linguistic homogenization, which means the gradual erosion of diverse language and cultural expressions due to pressures to conform to dominant language norms (Rahmani & Karimi, 2025). In globalized and digital contexts, dominant languages like English increasingly overshadow minority and heritage languages, resulting in significant declines in linguistic diversity and cultural distinctiveness (Skutnabb-Kangas & May, 2017).

Large language models amplify this process by reinforcing standardized linguistic patterns. Sourati et al. (2025) demonstrate that texts revised by LLMs show diminished stylistic and cultural variation, as features reflecting non-dominant languages and individual expression are systematically integrated into dominant forms. Similarly, Milička et al. (2025) found through multidimensional analysis that LLM-generated texts exhibit reduced stylistic variation compared to human writing, with AI maintaining more consistent and thus more homogenized output across registers. This change limits the expressive capacity of communication and diminishes the cultural frameworks that communities use to interpret experiences (Lin et al., 2025; Kuteeva & Andersson, 2024).

Zeng and Yang (2024) contend that English hegemony, reinforced through AI systems predominantly trained on English data, marginalizes minority languages and epistemologies. Without targeted initiatives for linguistic pluralism, LLMs exacerbate English dominance, undermining cognitive and cultural diversity (Zeng & Yang, 2024; Li et al., 2024). Plum et al. (2025) argue LLMs lack "cultural reasoning," defined as the ability to recognize and adjust for culture-specific knowledge, values, and norms, which sustains stereotypes and ignores minority perspectives (Plum et al., 2025; Seth, 2025).

Heritage language speakers thus face a fundamental conflict: technological inclusion versus cultural identity preservation (Fenech-Borg et al., 2025). While LLMs possess cultural knowledge, they remain insensitive to cultural differences in practice, often requiring manual correction for appropriate adaptation (Singh et al., 2024; Tenzer et al., 2025).

4.3 Assimilation Pressure and Authenticity Concerns in Technology-Mediated Communication

Across the studies reviewed, there is consistent and accumulating evidence that newcomers encounter both overt and subtle pressures to adopt host-country communication norms, often at the expense of heritage language practices, idiomatic expressions, and culturally grounded ways of speaking (Alshihry, 2024; Karpava, 2024; Tenzer et al., 2025).

Research repeatedly shows that such pressures can prompt individuals to question whether their linguistic choices reflect their authentic identities or merely conform to institutional expectations to particularly in contexts where mastery of the dominant language is framed as essential for educational and professional inclusion (Migliarini, 2024; Marrone, 2017). Emerging work further indicates that when AI enters this landscape, institutional norms become intertwined with LLM-mediated communication tools, shaping not only what newcomers articulate but also how they understand themselves as they navigate between home and host cultures (García & Wei, 2014; Feng et al., 2025).

Authenticity is central to immigrants' language experiences: many wrestle with whether their speech or writing sounds like "themselves" or simply satisfies others' expectations (Alshihry, 2024; Karpava, 2021; Eerdemutu et al., 2024). LLMs intensify this tension. Experimental work shows that LLM outputs exhibit high semantic alignment with prompts but relatively low stylistic alignment, prioritizing content over individual style (Durandard et al., 2025). Studies of LLM-driven editing and lexical shifts similarly find that AI revisions preserve core meaning while converging on more standardized, high-prestige forms of expression (Lin et al., 2025; Milička et al., 2025). The result is language that reads as polished and professional but feels alien to its author.

For newcomers adapting to a new culture, this dynamic is especially consequential. AI-assisted communication may appear fluent and acceptable yet lack authenticity, creating psychological strain and a disconnect between communication and self (Bélanger & Verkuyten, 2023; Karpava, 2024). At precisely the time when immigrants need to maintain a connection to their heritage identity while developing competence in the dominant language, LLM-mediated revision can erode that link by stripping linguistic markers of identity and reinforcing assimilation pressures. In doing so, it risks undermining the very adaptation and inclusion processes it purports to support (Alshihry, 2024; Migliarini, 2024).

Section V: LLMs in Context—Distinguishing Tools and Understanding Mechanisms

5.1 Large Language Models versus Neural Machine Translation: A Critical Distinction

To understand the challenges LLMs pose in cross-cultural communication, they must be distinguished from neural machine translation (NMT) systems. NMT systems translate text between languages, seeking to preserve meaning accurately. Using encoder–decoder architectures, they learn relationships between language pairs and perform direct translations such as "buenos días" → "good morning" (Ye, 2025; Boukhari & Regedor, 2025). LLMs, by contrast, generate fluent, contextually appropriate responses. They learn statistical patterns across languages that incorporate cultural knowledge and communicative norms (Hu et al., 2024; Li et al., 2024; Liu et al., 2024; Sun et al. 2025). When answering cultural questions, they synthesize information from training data, which may embed bias. A newcomer using Google Translate to read a sign relies on a tool with predictable limits. Asking ChatGPT for cultural advice, however, engages a system that seems informed yet often reflects Western perspectives and lacks the nuanced understanding users expect.

5.2 Authenticity and Naturalness Concerns in AI-Mediated Communication

Authenticity in AI-mediated cross-cultural communication involves not only identity preservation but also questions of genuine understanding and interaction. These concerns arise at several levels, including information about culture, AI-assisted communication, and relationships mediated through technology.

A growing body of research shows that LLMs often rely on surface-level stereotypes rather than genuine understanding of cultural values (Kharchenko et al., 2025; Lawton & Ibarrola, 2023). They can produce fluent descriptions of cultural practices yet frequently reflect Western interpretations rather than local perspectives. For instance, when describing Japanese notions of honor or Mexican family structures, LLMs depend mainly on English-language sources written by Western observers instead of knowledge from within those cultures. As a result, they appear culturally knowledgeable but reproduce secondhand knowledge about cultures rather than from them. For newcomers trying to understand a host culture, this distinction is crucial because such information may be accurate yet still miss the nuance and lived complexity of real practice.

A related issue is naturalness in communication. LLM-assisted writing may appear fluent and standardized yet feel less authentic to the writer's own voice and cultural background (Hwang et al., 2025). This tension between fluency and authenticity continues to challenge newcomers seeking ways to communicate that are both effective and true to self.

Section VI: Institutional and Policy Responses—From Prohibition to Participation

6.1 Educational Institution Responses: Between Prohibition and Integration

It's well documented that educational institutions have adopted diverse strategies for using LLMs in cross-cultural contexts, shaped by institutional goals, cultural values, and teaching philosophies. Approaches range from outright bans to structured integration frameworks that balance AI's benefits with the need for autonomy and academic integrity (Barnes et al., 2024; Cotton et al., 2024; Gulumbe et al., 2025; Nnorom, 2025).

The prohibition approach addresses valid concerns about integrity, dependency, and skill loss. However, it is difficult to enforce and can disadvantage non-native speakers who rely on AI to achieve native-level writing (Yusuf et al., 2024). Such bans can harm students who need assistance without deterring those who use AI covertly.

More progressive institutions adopt disclosure frameworks requiring students to acknowledge AI use while ensuring equitable access across cultures and socioeconomic groups (Yusuf et al., 2024). These frameworks accept AI's ubiquity and shift the question from *whether* students will use it to *how*: either to support learning or replace cognitive engagement.

Nordic universities base AI policies on values of trust, transparency, and openness (Butt, 2024; Cannavale et al., 2025; Masso et al., 2024; Rekman, 2024). Cultural contexts shape these designs, emphasizing collaboration over punishment and autonomy over control. Elsewhere, institutions prioritize innovation, competitiveness, or cost efficiency, producing divergent policies despite shared technological challenges (Cai & Yin, 2025; Goffi & Momcilovic, 2022; Han et al., 2025; Hongladarom & Bandasak, 2024; Kochupillai et al., 2022; Kum et al., 2024; Núñez,

2025; Popa Tache & Vâlcu, 2025; Wong, 2025).

Research highlights the importance of culturally responsive policies and sustained faculty development (Al-Zahrani & Alasmari, 2024; Ahmed, 2024). Effective AI integration requires solid infrastructure, faculty training, institutional support, and commitment to educational quality and cultural equity. Institutions that invest in comprehensive faculty programs more effectively distinguish between AI uses that enhance learning and those that undermine integrity or reinforce bias against non-native and marginalized students (Ma et al., 2024).

6.2 Community Organization Strategies: Building Support Beyond Institutions

Community organizations address challenges in AI-mediated communication by developing guidelines for responsible, culturally aware AI use, building peer support networks to reduce dependence, and offering education on both benefits and risks (Salas-Pilco et al., 2022). These initiatives promote digital inclusivity and cultural literacy, helping immigrants, international students, and minorities navigate an increasingly AI-mediated world.

Research on community interventions highlights the need for multidimensional solutions that combine educational, technological, and social approaches (Salas-Pilco et al., 2022; Marko et al., 2025). Technology alone is insufficient. Effective interventions focus on three dimensions: pedagogical (media literacy and critical AI evaluation), technological (accessible, multilingual, culturally responsive tools), and sociocultural (addressing power dynamics, cultural hierarchies, and identity issues in AI use).

Peer support programs are key to reducing dependency. They address the psychological and social dimensions institutional policies often overlook by promoting group reflection, setting boundaries between AI assistance and independent learning, and providing emotional support for those concerned about identity in AI-mediated communication (Salas-Pilco et al., 2022). Community organizations also act as intermediaries between policymakers and local populations, advocating for solutions tailored to each community's needs rather than one-size-fits-all governance models.

6.3 Professional and Organizational Adaptations: Navigating Workplace Complexities

Employers are developing AI workplace policies that recognize the technology's ubiquity while upholding expectations for cross-cultural competence, authentic communication, and ethical conduct (Tang et al., 2023; Rakova et al., 2020). Research shows that organizational culture, leadership support, and continuous training are essential for AI integration that strengthens work quality and inclusion (Ahmad et al., 2023; Einola & Khoreva, 2022).

A major challenge lies in distinguishing productive AI use from harmful practices. The workslop phenomenon, which indicate low-quality AI-generated content, remains difficult to regulate (Rakova et al., 2020; Bankins et al., 2024). Sharing such content, especially across cultural and linguistic boundaries, increases colleagues' cognitive load, erodes trust, and risks perpetuating stereotypes. Many organizations still lack frameworks to differentiate between beneficial AI tools, such as translation and accessibility support, and workslop that adds little value (Bankins et al., 2024).

Effective responses require AI literacy that merges technical and cultural competence

(Sienkiewicz-Małyjurek & Zyzak, 2024). Training should go beyond technical skills to include critical reflection on culturally sensitive AI use, the preservation of nuance, and transparent communication across cultures. Organizations that foster supportive cultures, model ethical AI use, and sustain employee development achieve greater success in using AI to enhance rather than diminish cross-cultural communication (Sienkiewicz-Małyjurek & Zyzak, 2024).

6.4 Cross-Cultural and Policy Considerations: Values, Disparities, and Governance

Global disparities in AI adoption and digital infrastructure undermine equitable cross-cultural communication. Studies reveal stark divides between high- and low-resource regions, where limited access, poor service quality, and weak institutional capacity hinder context-specific AI governance (Al-Zahrani & Alasmari, 2024; Ahmed, 2024; Marko et al., 2025). As a result, marginalized groups, including speakers of low-resource languages, residents of developing regions, immigrants, and refugees, struggle to access supportive AI tools while being overexposed to Western-trained systems that reinforce cultural bias (Ahmed, 2024). The compounding effect is a form of technological marginalization that mirrors and intensifies existing global inequalities.

Addressing these challenges requires context-specific policies attuned to local realities rather than universal models. Effective strategies must span four dimensions: technical (culturally responsive, accessible tools), governance (locally grounded frameworks), educational (digital literacy and critical AI skills), and equity (reducing global disparities in access and quality) (Abbasnejad et al., 2025; Abbasi et al., 2025; Kudriashova & Martynenko, 2025).

Discussion: Implications, Limitations, and Paths Forward

7.1 Synthesis of Major Findings

Taken together, the results show that cultural bias in large language models is not incidental but structural. At the architectural level, models are trained on corpora that overrepresent English and Western epistemologies, encoding Western-centric value priorities into their parameters and alignment processes even when outputs appear neutral. At the experiential level, newcomers encounter this bias as cultural misrepresentation, stereotype reinforcement, and communication-style mismatch: LLMs often recast culturally grounded reasoning into Western academic forms, mishandle idioms and metaphors, and recommend interaction norms misaligned with high-context or relational cultures.

These patterns translate into what the review terms an equity paradox. Non-native speakers face a cruel double bind in AI detection systems, where both authentic writing and AI-assisted improvement can trigger suspicion, while speakers of low-resource languages confront higher costs and lower-quality service. Workplace uses of LLMs can generate workslop that erodes trust and disproportionately harms those who depend on AI for linguistic support. In education, performance gaps, brittleness across languages, and misaligned benchmarks further disadvantage multilingual learners.

The findings also document processes of identity erasure. LLM-based rewriting preserves semantic content but systematically strips stylistic and cultural markers, pushing users toward standardized global norms and weakening heritage language vitality. For newcomers, this

produces a tension between fluency and authenticity, as AI-assisted communication may read as acceptable yet feel detached from their sense of self. Distinguishing LLMs from more bounded tools like neural machine translation clarifies why these effects arise: LLMs do not simply translate but synthesize and normalize cultural knowledge.

Finally, the review shows that institutional and policy responses remain uneven. Educational, community, and workplace initiatives are beginning to address AI's role in cross-cultural communication, but many frameworks remain top-down and insufficiently participatory. Across studies, a consistent message emerges: mitigating harm and supporting equitable use will require governance, design, and community practices that explicitly center underrepresented users rather than treating LLMs as culturally neutral tools.

7.2 Theoretical Implications

LLMs extend identity-protective cognition to the technological sphere. Users interpret information through the lens of their group identity rather than rational evaluation; algorithms also embed group-centric values during training and deployment. LLMs amplify majority-culture patterns, creating asymmetries that advantage majority users while marginalizing others. Simultaneously, LLMs redefine authenticity, privileging communication styles aligned with dominant training data rather than with lived cultural experience.

7.3 Implications for Newcomers

Newcomers and immigrants face a compound burden: underrepresentation of their cultures in LLMs, inadequate service at high costs, and homogenization of expression that threatens cultural preservation. For vulnerable populations, LLMs pose risks with opportunities.

7.4 Policy and Governance Implications

Current governance frameworks lack participatory structures centered on affected communities. Effective governance requires treating AI as a question of power and representation, fundamentally political issues demanding democratic rather than purely technical solutions. Governance must enable data sovereignty and give underrepresented communities control over how their cultures are represented, ensuring collaboration rather than imposition.

7.5 Technical and Design Implications

Corpus curation must intentionally represent diverse cultural values and non-Western epistemologies. Community participation and data sovereignty are prerequisites, not optional. Evaluation must extend beyond technical accuracy to assess cultural appropriateness and the impact on equity, utilizing culturally diverse evaluators. Architectures must integrate transparency and community feedback rather than remain opaque black boxes.

7.6 Limitations and Future Research

This semi-systematic review strikes a balance between rigor and feasibility in a rapidly evolving field. Limitations include reliance on English-language sources, the

2020–2025 timeframe, limited longitudinal evidence on cultural identity formation, and underrepresentation of research from non-English-speaking regions and lower-income countries. Future research should examine identity development through longitudinal cross-cultural studies, evaluate community-based and technical interventions, and analyze how LLM bias intersects with other forms of marginalization.

7.7 Conclusion: Toward Technology that Serves Rather Than Dominates

LLM proliferation demands attention to equity, cultural preservation, and authentic understanding. These systems embed biases advantaging some groups while risking cultural erasure where diversity most needs protection. Our task is to determine which systems to build, whose interests they serve, and what values they express. AI governance is fundamentally about power and representation, requiring participatory structures that center on and represent affected communities.

LLMs cannot replace the human work of building cross-cultural understanding. For newcomers navigating integration, for communities protecting linguistic and cultural distinctiveness, and for those committed to equitable communication, the key question is not only what AI can do but what it should refrain from doing. The tools will remain; our task is to use them in ways that support, rather than displace, the slow and demanding work of genuine human connection.

References

Abbasi, B. N., Wu, Y., & Luo, Z. (2025). Exploring the impact of artificial intelligence on curriculum development in global higher education institutions. Education and Information Technologies, 30(1), 547-581.

Abbasnejad, B., Soltani, S., Taghizadeh, F., & Zare, A. (2025). Developing a multilevel framework for AI integration in technical and engineering higher education: insights from bibliometric analysis and ethnographic research. Interactive Technology and Smart Education.

Abdilla, A., & Crawford, K. (2020). Indigenous knowledge systems are not supplementary to Western paradigms. International Journal of Communication, 14, 1–12.

Agarwal, D., Naaman, M., & Vashistha, A. (2025, April). AI suggestions homogenize writing toward western styles and diminish cultural nuances. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (pp. 1-21).

Ahia, O., Kumar, S., Gonen, H., Kasai, J., Mortensen, D. R., Smith, N. A., & Tsvetkov, Y. (2023). Do all languages cost the same? tokenization in the era of commercial language models. arXiv preprint arXiv:2305.13707.

Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M., Irshad, M., Arraño-Muñoz, M., & Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. Humanities and Social Sciences Communications, 10(1), 1-14.

Ahmed, S. (2024). Beyond Human Teachers: Post-Humanist Perspectives on AI, Cultural Inequities, and Educational Transformation. Journal of Posthumanism, 4(3), 364-373.

Al-Zahrani, A. M., & Alasmari, T. M. (2024). Exploring the impact of artificial intelligence on higher education: The dynamics of ethical, social, and educational implications. Humanities and

Social Sciences Communications, 11(1), 1-12.

Algouzi, S., & Alzubi, A. A. F. (2023). The study of AI-mediated communication and socio-cultural language-related variables: Gmail reply suggestions. Applied Artificial Intelligence, 37(1), 2175114.

Allan, K., Azcona, J., Sripada, S., Leontidis, G., Sutherland, C. A., Phillips, L. H., & Martin, D. (2025). Stereotypical bias amplification and reversal in an experimental model of human interaction with generative artificial intelligence. Royal Society Open Science, 12(4), 241472.

Alshihry, M. A. (2024). Heritage language maintenance among immigrant youth: Factors influencing proficiency and identity. In S. Karpava (Ed.), Heritage language policy in early childhood education (pp. 45–65). Springer.

Alyafeai, Z., Alshaibani, M. S., AlKhamissi, B., Luqman, H., Alareqi, E., & Fadel, A. (2023). Taqyim: Evaluating arabic nlp tasks using chatgpt models. arXiv preprint arXiv:2306.16322.

Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., … & Veríssimo, D. (2023). The manifold costs of being a non-native English speaker in science. PLoS biology, 21(7), e3002184.

Arıcı, F. (2024). Examination of research conducted on the use of artificial intelligence in science education. Sakarya University Journal of Education, 14(3), 539-562.

Asfahani, A. M. (2022). The impact of artificial intelligence on industrial-organizational psychology: A systematic review. The Journal of Behavioral Science, 17(3), 125-139.

Ashraf, Y., Wang, Y., Gu, B., Nakov, P., & Baldwin, T. (2025, April). Arabic dataset for LLM safeguard evaluation. In Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 5529-5546).

Asseri, B., Abdelaziz, E., & Al-Wabil, A. (2025). Prompt Engineering Techniques for Mitigating Cultural Bias Against Arabs and Muslims in Large Language Models: A Systematic Review. arXiv preprint arXiv:2506.18199.

Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2024). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. Journal of organizational behavior, 45(2), 159-182.

Bao, T., Zhao, Y., Mao, J., & Zhang, C. (2025). Examining linguistic shifts in academic writing before and after the launch of ChatGPT: a study on preprint papers. Scientometrics, 1-31.

Bareis, J., & Katzenbach, C. (2022). Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics. Science, Technology, & Human Values, 47(5), 855-881.

Barnes, A. J., Zhang, Y., & Valenzuela, A. (2024). AI and culture: Culturally dependent responses to AI systems. Current Opinion in Psychology, 58, 101838.

Bélanger, E., & Verkuyten, M. (2023). Language and belonging among older immigrants. Journal of Ethnic and Migration Studies, 49(16), 3973–3992. https://doi.org/10.1080/1369183X.2023.2280994

Bignotti, F. (2025). Potential bias in AI: cultural representation and the marginalization of African art.

Binkyte, R. (2025). Interactional Fairness in LLM Multi-Agent Systems: An Evaluation Framework. arXiv preprint arXiv:2505.12001.

Birhane, A., Isaac, W., Prabhakaran, V., Diaz, M., Elish, M. C., Gabriel, I., & Mohamed, S. (2022, October). Power to the people? Opportunities and challenges for participatory AI. In Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (pp. 1-8).

Björk, B. C., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. Journal of informetrics, 7(4), 914-923.

Borah, A., & Mihalcea, R. (2024). Towards implicit bias detection and mitigation in multi-agent llm interactions. arXiv preprint arXiv:2410.02584.

Boukhari, O., & Regedor, A. M. (2025). Language as a Barrier, Bridge and Battleground: A Study of Linguistic Experiences in the Context of Moroccan Migration to Southern Portugal. International Journal of Linguistics, Literature and Translation, 8(5), 727-742.

Brandt, B. (2023). Replicating human bias through synthetic data generation using deep learning (Doctoral dissertation, University of Wisconsin-Whitewater).

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. The Quarterly Journal of Economics, 140(2), 889-942.

Butt, J. S. (2024). A comparative study about the use of artificial intelligence (AI) in public administration of Nordic states with other European economic sectors. Euro Economica, 43(1), 40-66.

Cai, C., & Yin, J. (2025). Cultural and ethical foundations of AI governance divergence: a comparative analysis of China and the west. Revista Política Internacional, 7(1), 215-233.

Cannavale, C., Claudio, L., & Koroleva, D. (2025). Digitalisation and artificial intelligence development. A cross-country analysis. European Journal of Innovation Management, 28(11), 112-130.

Chatterji, A., Cunningham, T., Deming, D. J., Hitzig, Z., Ong, C., Shan, C. Y., & Wadman, K. (2025). How people use chatgpt (No. w34255). National Bureau of Economic Research.

Chiu, Y. Y., Jiang, L., Lin, B. Y., Park, C. Y., Li, S. S., Ravi, S., Bhatia, M., Antoniak, M., Tsvetkov, Y., Shwartz, V., & Choi, Y. (2024). CulturalBench: A robust, diverse and challenging benchmark on measuring the (lack of) cultural knowledge of LLMs. arXiv preprint. https://arxiv.org/abs/2410.02677

Cook, D. J., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: synthesis of best evidence for clinical decisions. Annals of internal medicine, 126(5), 376-380.

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in education and teaching international, 61(2), 228-239.

Crystal, D. (2003). English as a global language. Cambridge university press.

Dairo, D. Navigating Cultural Diversity in the Digital Age: Legal Perspectives on Opportunities and Challenges of Artificial Intelligence. In Cultural Odyssey: 20 Years of Implementation of UNESCO's 2005 Convention in Nigeria (p. 181).

Darginavičienė, I. (2023). The multilingualism: Language and cultural identity. LOGOS-A Journal of Religion, Philosophy, Comparative Cultural Studies and Art, (116), 167-174.

De Sousa Santos, B. (2014). Epistemologies of the South: Justice against epistemicide. Paradigm Publishers.

Durandard, N., Dhawan, S., & Poibeau, T. (2025). LLMs stick to the point, humans to style: Semantic and stylistic alignment in human and LLM communication. Proceedings of the 26th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 206–213.

Durmus, E., Nguyen, K., Liao, T. I., Schiefer, N., Askell, A., Bakhtin, A., … & Ganguli, D. (2023). Towards measuring the representation of subjective global opinions in language models. arXiv preprint arXiv:2306.16388.

Eger, S., Cao, Y., D'Souza, J., Geiger, A., Greisinger, C., Gross, S., … & Miller, T. (2025). Transforming science with large language models: A survey on ai-assisted scientific discovery, experimentation, content generation, and evaluation. arXiv preprint arXiv:2502.05151.

Ehrensberger-Dow, M., Albl-Mikasa, M., Andermatt, K., Hunziker Heeb, A., & Lehr, C. (2020). Cognitive load in processing ELF: Translators, interpreters, and other multilinguals. Journal of English as a lingua franca, 9(2), 217-238.

Einola, K., & Khoreva, V. (2023). Best friend or broken tool? Exploring the co-existence of humans and artificial intelligence in the workplace ecosystem. Human Resource Management, 62(1), 117-135.

Elgamal, A. M. A. (2019). Cognitive Factors Affecting Language Learning and Acquisition of Native and Non-Native Speakers. Journal of Research in Curriculum Instruction and Educational Technology, 4(4), 135-152.

Fenech-Borg, E. Z., Meznaric-Kos, T. P., Lekovic-Bojovic, M. D., & Hentze-Djurhuus, A. J. (2025). The Cultural Gene of Large Language Models: A Study on the Impact of Cross-Corpus Training on Model Values and Biases. arXiv preprint arXiv:2508.12411.

Feng, H., Li, K., & Zhang, L. J. (2025). What does AI bring to second language writing? A systematic review (2014-2024). Language Learning & Technology, 29(1).

Ferdaus, M. M., Abdelguerfi, M., Loup, E., N. Niles, K., Pathak, K., & Sloan, S. (2024). Towards trustworthy ai: A review of ethical and robust large language models. ACM Computing Surveys.

Ferré, P., Fraga, I., & Hinojosa, J. A. (2025). The interplay between language and emotion: a narrative review. Cognition and Emotion, 1-28.

Fielding, R. (2021). A multilingual identity approach to intercultural stance in language learning. The Language Learning Journal, 49(4), 466-482.

Fock, A., & Siller, H. S. (2025). Generative Artificial Intelligence in Secondary STEM Education in the Light of Human Flourishing: A Scoping Literature Review.

Fraser, K. C., Dawkins, H., & Kiritchenko, S. (2025). Detecting ai-generated text: Factors influencing detectability with current methods. Journal of Artificial Intelligence Research, 82, 2233-2278.

Gao, F. (2021). Negotiation of native linguistic ideology and cultural identities in English learning: a cultural schema perspective. Journal of Multilingual and Multicultural Development, 42(6), 551-564.

Garcia, M. B. (2025). ChatGPT as an Academic Writing Tool: Factors Influencing Researchers' Intention to Write Manuscripts Using Generative Artificial Intelligence. International Journal of Human–Computer Interaction, 1-15.

García, O., & Wei, L. (2014). Translanguaging: Language, bilingualism and education. Palgrave Macmillan.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. Proceedings of the National Academy of Sciences, 115(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Ghimire, A. (2025). Utilizing ChatGPT to integrate world English and diverse knowledge: A transnational perspective in critical artificial intelligence (AI) literacy. Computers and Composition, 75, 102913.

Glickman, M., & Sharot, T. (2025). How human–AI feedback loops alter human perceptual, emotional and social judgements. Nature Human Behaviour, 9(2), 345-359.

Goffi, E., & Momcilovic, A. (2022). Respecting cultural diversity in ethics applied to AI: A new approach for a multicultural governance. Misión Jurídica, 15(23), 111-122.

Gopalakrishnan, S., & Ganeshkumar, P. (2013). Systematic reviews and meta-analysis: understanding the best evidence in primary healthcare. Journal of family medicine and primary care, 2(1), 9-14.

Gulumbe, B. H., Audu, S. M., & Hashim, A. M. (2025). Balancing AI and academic integrity: What are the positions of academic publishers and universities?. AI & SOCIETY, 40(3), 1775-1784.

Guo, J., & Xu, Y. (2025). Your AI Bosses Are Still Prejudiced: The Emergence of Stereotypes in LLM-Based Multi-Agent Systems. arXiv preprint arXiv:2508.19919.

Gwagwa, A., Kraemer-Mbula, E., Rizk, N., Rutenberg, I., & de Beer, J. (2020). Artificial intelligence (AI) deployments in Africa: Benefits, challenges and policy dimensions. The African Journal of Information and Communication, 26, 1–28. https://doi.org/10.23962/10539/30361

Haase, J., & Pokutta, S. (2025). Beyond Static Responses: Multi-Agent LLM Systems as a New Paradigm for Social Science Research. arXiv preprint arXiv:2506.01839.

Han, X., HL Kaas, M., & Wang, C. D. (2025). A cross-cultural examination of fairness beliefs in human-AI interaction. Han, X., Kass, M., & Wang, C.(forthcoming). A cross-cultural examination of fairness beliefs in human-AI interaction. In Ethics of Institutional Beliefs: From Theoretical to Empirical. Edward Elgar Publishing.

Hamann, S. A., Giang, J., De Maeseneer, M. G., Nijsten, T. E., & van den Bos, R. R. (2017). Editor's choice–five year results of great saphenous vein treatment: a meta-analysis. *European journal of vascular and endovascular surgery*, *54*(6), 760-770.

Havaldar, S., Cho, Y. M., Rai, S., & Ungar, L. (2025, November). Culturally-Aware Conversations: A Framework & Benchmark for LLMs. In Proceedings of the Fourth Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCI+ NLP) (pp. 220-229).

Hergert, L., Berend, G., Szegedy, M., Turan, G., & Jelasity, M. (2025). On the Brittleness of LLMs: A Journey around Set Membership. arXiv preprint arXiv:2511.12728.

Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). AI generates covertly racist decisions about

people based on their dialect. Nature. https://doi.org/10.1038/s41586-024-07856-5

Hongladarom, S., & Bandasak, J. (2024). Non-western AI ethics guidelines: Implications for intercultural ethics of technology. Ai & Society, 39(4), 2019-2032.

Hoppers, O. C. (2002). Indigenous knowledge and the integration of knowledge systems. Indigenous knowledge and the integration of knowledge systems: Towards a philosophy of articulation, 2-22.

Hsiao, C. Y. (2021). Online and Offline Adaptation among Transnational Newcomers: Technology-mediated Social Exchange and Trust Development (Doctoral dissertation).

Hu, H., Zhou, Q., & Hashim, H. (2025). Negotiating identity in the age of ChatGPT: non-native English researchers' experiences with AI-assisted academic writing. Humanities and Social Sciences Communications, 12(1), 1-11.

Hu, X., Zhang, F., Chen, S., & Yang, Z. (2024). Unveiling the statistical foundations of chain-of-thought prompting methods. arXiv preprint arXiv:2408.14511.

Hwang, A. H. C., Liao, Q. V., Blodgett, S. L., Olteanu, A., & Trischler, A. (2025). 'It was 80% me, 20% AI': Seeking Authenticity in Co-Writing with Large Language Models. Proceedings of the ACM on Human-Computer Interaction, 9(2), 1-41.

Jiang, Y., Hao, J., Fauss, M., & Li, C. (2024). Detecting ChatGPT-generated essays in a large-scale writing assessment: Is there a bias against non-native English speakers?. Computers & Education, 217, 105070.

Jiang, Y., Zhao, J., Yuan, Y., Zhang, T., Huang, Y., Zhang, Y., … & Li, X. (2025). Never compromise to vulnerabilities: A comprehensive survey on ai governance. arXiv preprint arXiv:2508.08789.

Joubert, M., & Sibanda, B. (2022). Whose language is it anyway? Students' sense of belonging and role of English for higher education in the multilingual, South African context. South African Journal of Higher Education, 36(6), 47-66.

Kamran, F. (2024). Decolonizing artificial intelligence: Indigenous knowledge and digital epistemic justice. Journal of Critical AI Studies, 2(1), 45–68.

Karpava, S. (2021). Heritage language use, maintenance and transmission by second-generation immigrants in Cyprus. In P. Romanowski & M. Jedynak (Eds.), Current Research in Bilingualism and Education (pp. 83–108). Springer.

Karpava, S. (2024). The hybrid linguistic and cultural identity of second-generation immigrants in Cyprus. In S. Karpava (Ed.), Hybrid linguistic and cultural identities in migration contexts (pp. 101–124). Routledge.

Keleg, A. (2025). LLM Alignment for the Arabs: A Homogenous Culture or Diverse Ones?. arXiv preprint arXiv:2503.15003.

Keller, J. D., & Potthast, R. (2024). AI-based data assimilation: Learning the functional of analysis estimation. arXiv preprint arXiv:2406.00390.

Khan, R., Qamar, M. T., Ansari, M. S., & Yasmeen, J. (2025). Enhancing or impairing? Exploring Indian EFL learners' academic writing narratives with ChatGPT. Cogent Education, 12(1), 2514329.

Kharchenko, J., Roosta, T., Chadha, A., & Shah, C. (2024). How well do llms represent values across cultures? empirical analysis of llm responses based on hofstede cultural dimensions. arXiv preprint arXiv:2406.14805.

Kiramba, L. K., & Oloo, J. A. (2023). "It's OK. She doesn't even speak English": Narratives of language, culture, and identity negotiation by immigrant high school students. Urban Education, 58(3), 398-426.

Klassen, T. P., Lawson, M. L., & Moher, D. (2005). Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. Journal of clinical epidemiology, 58(8), 769-776.

Kochupillai, M., Kahl, M., Schmitt, M., Taubenböck, H., & Zhu, X. X. (2022). Earth observation and artificial intelligence: Understanding emerging ethical issues and opportunities. IEEE Geoscience and Remote Sensing Magazine, 10(4), 90-124.

Koo, W. W. (2025). Cross-lingual effects of AI-generated content on human work. Scientific Reports, 15(1), 30949.

Kum, H. C., Bedrick, S., & Weigle, M. C. (2024). Challenges in Data Science. Digital Ethology: Human Behavior in Geospatial Context, 33, 211.

Kuteeva, M., & Andersson, M. (2024). Diversity and Standards in Writing for Publication in the Age of AI—Between a Rock and a Hard Place. Applied Linguistics, 45(3), 561-567.

Kwak, Y., & Pardos, Z. A. (2024). Bridging large language model disparities: Skill tagging of multilingual educational content. British Journal of Educational Technology, 55(5), 2039-2057.

Lawton, T., Grace, K., & Ibarrola, F. J. (2023, July). When is a tool a tool? user perceptions of system agency in human–ai co-creative drawing. In Proceedings of the 2023 ACM Designing Interactive Systems Conference (pp. 1978-1996).

Lee, E., & Moshirnia, A. (2024). The AI Penalty: Is There a Bias against AI-Generated Works?. Mich. St. L. Rev., 641.

Lege, R. P. Auditing the Fairness of AI-Detection Tools: A Comparative Study of ESL, Published, and AI-Generated Texts and Their Misclassification Risks. International Journal of Teaching, Learning and Education, 4(5), 638273.

Lehdonvirta, V. (2022). Cloud empires: How digital platforms are overtaking the state and how we can regain control. Mit press.

Li, C., Chen, M., Wang, J., Sitaram, S., & Xie, X. (2024). Culturellm: Incorporating cultural differences into large language models. Advances in Neural Information Processing Systems, 37, 84799-84838.

Li, J., Papay, S., & Klinger, R. (2025). Are Humans as Brittle as Large Language Models?. arXiv preprint arXiv:2509.07869.

Li, Y., Huang, Y., Wang, H., Cheng, Y., Zhang, X., Zou, J., & Sun, L. Evaluating Large Language Models with Psychometrics. In Large Language Models for Scientific and Societal Advances.

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. Patterns, 4(7).

Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., … & Zou, J. (2025). Quantifying large language model usage in scientific papers. Nature Human Behaviour, 1-11.

Lin, D., Zhao, N., Tian, D., & Li, J. (2025). ChatGPT as Linguistic Equalizer? Quantifying LLM-Driven Lexical Shifts in Academic Writing. arXiv preprint arXiv:2504.12317.

Lin, Z., & Zhao, X. Cultural Tailoring Paradox: Navigating Perceived Homophily and AI Bias in Generative AI-Mediated Communication Among Black Communities. Available at SSRN 5261067.

Liu, X., Wu, Z., Wu, X., Lu, P., Chang, K. W., & Feng, Y. (2024). Are llms capable of data-based statistical and causal reasoning? benchmarking advanced quantitative reasoning with data. arXiv preprint arXiv:2402.17644.

Liu, Z., Zhang, J., Jiang, H., You, W., Pan, Y., Xu, S., … & Liu, T. (2025). AImanities and Mirror of Collectivized Mind: Philosophy Theories of Large Language Models.

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. Smart Learning Environments, 11(1), 50.

Lodoen, S., & Orchard, A. (2025). Ethics and Persuasion in Reinforcement Learning from Human Feedback: A Procedural Rhetorical Approach. arXiv preprint arXiv:2505.09576.

Lyu, Y., & Du, Y. (2025). The ethical evaluation of large language models and its optimization. AI and Ethics, 1-14.

Ma, D., Akram, H., & Chen, I. H. (2024). Artificial intelligence in higher education: A cross-cultural examination of students' behavioral intentions and attitudes. International Review of Research in Open and Distributed Learning, 25(3), 134-157.

Mao, R., Liu, Q., Li, X., Cambria, E., & Hussain, A. (2025). Bridging Minds and Machines: Toward an Integration of AI and Cognitive Science. arXiv preprint arXiv:2508.20674.

Marko, J. G. O., Neagu, C. D., & Anand, P. B. (2025). Examining inclusivity: the use of AI and diverse populations in health and social care: a systematic review. BMC Medical Informatics and Decision Making, 25(1), 57.

Marrone, G. (2017). Linguistic and cultural assimilation as a human capital process. IZA Journal of Migration, 6(1), 1–27.

Masso, A., Kaun, A., & Van Noordt, C. (2024). Basic values in artificial intelligence: comparative factor analysis in Estonia, Germany, and Sweden. AI & society, 39(6), 2775-2790.

Mehdizadeh, A., & Hilbert, M. (2025). Homophily-induced emergence of biased structures in LLM-based multi-agent AI systems. Social Network Analysis and Mining, 15(1), 1-25.

Migliarini, V. (2024). Performing the good (im)migrant: Inclusion and expectations of (dis)abled asylum-seeking students. International Journal of Inclusive Education, 28(12), 1432–1451.

Milička, J., Marklová, A., & Cvrček, V. (2025). Benchmark of stylistic variation in LLM-generated texts. Corpus Linguistics and Linguistic Theory, 21(2), 255–283.

Naous, T., Laban, P., Xu, W., & Neville, J. (2025). Flipping the Dialogue: Training and Evaluating User Language Models. *arXiv preprint arXiv:2510.06552*.

Nguyen, J. K. (2024). Human bias in AI models? Anchoring effects and mitigation strategies in large language models. Journal of Behavioral and Experimental Finance, 43, 100971. https://doi.org/10.1016/j.jbef.2024.100971

Nguyen, X.-P., Aljunied, M., Joty, S., & Bing, L. (2024). Democratizing LLMs for low-resource languages by leveraging their English dominant abilities with linguistically-diverse prompts. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 3501–3516. https://doi.org/10.18653/v1/2024.acl-long.192

Niederhoffer, K., Kellerman, G. R., Lee, A., Liebscher, A., Rapuano, K., & Hancock, J. T. (2025). AI-Generated "Workslop" Is Destroying Productivity. Harvard Business Review.

Nnorom, I. C. (2025). Ethical Considerations in Artificial Intelligence and Academic Integrity: Balancing Technology and Human Values. AI and Ethics, Academic Integrity and the Future of Quality Assurance in Higher Education, 15.

Núñez, S. H. (2025). Technology Transfer to Latin American Countries: Drifting Away from the United States and China?. Taylor & Francis.

Ozbek-Damar, S. (2025). Second Language (L2)–Speaker Immigrant Women's Perspectives on Identity Construction and L2 Socialization in a Community in the Southwest US(Doctoral dissertation, Arizona State University).

Öztürk, D. (2021). What does artificial intelligence mean for organizations? A systematic review of organization studies research and a way forward. The Impact of Artificial Intelligence on Governance, Economics and Finance, Volume I, 265-289.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. bmj, 372.

Pakray, P., Gelbukh, A., & Bandyopadhyay, S. (2025). Natural language processing applications for low-resource languages. Natural Language Processing, 31(2), 183-197.

Pareek, A. (2025). A Multi-Component AI Framework for Computational Psychology: From Robust Predictive Modeling to Deployed Generative Dialogue. arXiv preprint arXiv:2510.21720.

Pedersen, B. S., Sørensen, N., Nimb, S., Hansen, D. H., Olsen, S., & Al-Laith, A. (2025, March). Evaluating llm-generated explanations of metaphors–a culture-sensitive study of danish. In Proceedings of the Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies (NoDaLiDa/Baltic-HLT 2025) (pp. 470-479).

Pérez Torres, M., Couso Lagarón, D., & Marquez Bargalló, C. (2023). Evaluation of STEAM project-based learning (STEAM PBL) instructional designs from the STEM practices perspective. *Education Sciences*, *14*(1), 53.

Peters, U., & Carman, M. (2024). Cultural bias in explainable AI research: A systematic analysis. Journal of Artificial Intelligence Research, 79, 971-1000.

Petrov, A., La Malfa, E., Torr, P., & Bibi, A. (2023). Language model tokenizers introduce unfairness between languages. Advances in neural information processing systems, 36, 36963-36990.

Plum, A., Lutgen, A. M., Purschke, C., & Rettinger, A. (2025). Identity-Aware Large Language Models require Cultural Reasoning. arXiv preprint arXiv:2510.18510.

Popa Tache, C. E., & Vâlcu, E. N. (2025). Artificial Intelligence and Corporate Liability Towards a New Legal-Ethical Contract in the Dynamics of Emerging Global Human Rights Convergences. Jurid. Trib.-Rev. Compar. &Int'l L., 15, 281.

Prakash, A., Aggarwal, S., Varghese, J. J., & Varghese, J. J. (2025). Writing without borders: AI and cross-cultural convergence in academic writing quality. Humanities and Social Sciences

Communications, 12(1), 1-11.

Qadri, R., Diaz, M., Wang, D., & Madaio, M. (2025). The case for" thick evaluations" of cultural representation in ai. arXiv preprint arXiv:2503.19075.

Rahman, A., Bowlin, G., Mohanty, B., & McGunigal, S. (2024). Towards Linguistically-Aware and Language-Independent Tokenization for Large Language Models (LLMs). arXiv preprint arXiv:2410.03568.

Rakova, B., Yang, J., Cramer, H., & Chowdhury, R. (2021). Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), 1-23.

Rao, P. S. (2019). The role of English as a global language. Research journal of English, 4(1), 65-79.

Rekman, O. (2024). Nordic Ethical AI Expert Group: Policy Recommendations for Ethical and Responsible AI.

Reusens, M., Kopetzky, D., & Buitelaar, P. (2024). Anchoring effects in LLM response quality: Systematic performance degradation for non-native English speakers. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (pp. 1–15).

Richardson, N., & Antonello, M. (2022). People at work 2022: A global workforce view. ADP Research Institute, 17, 2023.

Robertson, C., & Wolff, P. (2025). Llm world models are mental: Output layer evidence of brittle world model use in llm mechanical reasoning. arXiv preprint arXiv:2507.15521.

Rodrigues, L., Xavier, C., Costa, N., Batista, H., Silva, L. F. B., Chaleghi de Melo, W., … & Ferreira Mello, R. (2025, March). LLMs Performance in Answering Educational Questions in Brazilian Portuguese: A Preliminary Analysis on LLMs Potential to Support Diverse Educational Needs. In Proceedings of the 15th International Learning Analytics and Knowledge Conference (pp. 865-871).

Sætra, H. S. (2024). Preprinting in AI Ethics: Towards a Set of Community Guidelines. Available at SSRN 4598223.

Sahebi, S., & Formosa, P. (2025). The AI-mediated communication dilemma: epistemic trust, social media, and the challenge of generative artificial intelligence. Synthese, 205(3), 1-24.

Salas-Pilco, S. Z., Xiao, K., & Oshima, J. (2022). Artificial intelligence and new technologies in inclusive education for minority students: A systematic review. Sustainability, 14(20), 13572.

Sallam, M., Al-Adwan, A. S., Mijwil, M. M., Abdelaziz, D. H., Al-Qaisi, A., Ibrahim, O. M., & Sallam, M. (2025). Technology Readiness, Social Influence, and Anxiety as Predictors of University Educators' Perceptions of Generative AI Usefulness and Effectiveness.

Sallam, M., & Mousa, D. (2024). Evaluating ChatGPT performance in Arabic dialects: A comparative study showing defects in responding to Jordanian and Tunisian general health prompts. Mesopotamian Journal of Artificial Intelligence in Healthcare, 2024, 1-7.

Sanguinetti, P., & Palomo, B. (2025). Bridging Gaps in AI Representation: A Cross-Cultural Analysis of Media Imagery. Journalism Practice, 1-21.

Sato, K., Kaneko, H., & Fujimura, M. (2024). Reducing cultural hallucination in non-english languages via prompt engineering for large language models. OSF Preprints, 10.

Savadori, L., Dickson, M. M., Micciolo, R., & Espa, G. (2022). The polarizing impact of numeracy, economic literacy, and science literacy on the perception of immigration. Plos one, 17(10), e0274680.

Segerer, R. (2025). Cultural Value Alignment in Large Language Models: A Prompt-based Analysis of Schwartz Values in Gemini, ChatGPT, and DeepSeek. arXiv preprint arXiv:2505.17112.

Seth, A. (2025). Cultural Variability and Bias in Online Social Interactions and Large Language Models (Doctoral dissertation).

Shan, S., Li, Y., & Zhou, W. (2024). Cross-cultural implications of LLM deployment: Extended comparative analysis. International Journal of Human-Computer Studies, 185, 103198.

Shen, H., Clark, N., & Mitra, T. (2025, November). Mind the Value-Action Gap: Do LLMs Act in Alignment with Their Values?. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (pp. 3097-3118).

Singh, P., Patidar, M., & Vig, L. (2024). Translating across cultures: Llms for intralingual cultural adaptation. arXiv preprint arXiv:2406.14504.

Solatorio, A. V., Macalaba, R., & Liounis, J. (2025). Large Language Models and Synthetic Data for Monitoring Dataset Mentions in Research Papers. arXiv preprint arXiv:2502.10263.

Sourati, Z., Karimi-Malekabadi, F., Ozcan, M., McDaniel, C., Ziabari, A., Trager, J., … & Dehghani, M. (2025). The shrinking landscape of linguistic diversity in the age of large language models. arXiv preprint arXiv:2502.11266.

Su, J., Zhang, J., Ullrich, K., Bottou, L., & Ibrahim, M. (2025). A single character can make or break your LLM evals. arXiv preprint arXiv:2510.05152.

Suh, S., Bang, J., & Han, J. W. (2025). The Shift Towards Preprints in AI Policy Research: A Comparative Study of Preprint Trends in the US, Europe, and South Korea. arXiv preprint arXiv:2505.03835.

Sumartana, I. M., Hudiananingsih, P. D., & Rouf, M. A. (2025). Balancing globalization and linguistic heritage involves preserving mother tongues through inclusive education that values cultural identity and language diversity. Journal of Language, Literature, Social and Cultural Studies, 3(2), 179-196.

Sun, M., Han, R., Jiang, B., Qi, H., Sun, D., Yuan, Y., & Huang, J. (2025). A survey on large language model-based agents for statistics and data science. The American Statistician, 1-14.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of business research*, *104*, 333-339.

Tafa, T. O., Hashim, S. Z. M., Othman, M. S., Alhussian, H., Nasser, M., Abdulkadir, S. J., … & Bena, Y. A. (2025). Machine Translation Performance for LowResource Languages: A Systematic Literature Review. IEEE Access.

Tang, P., Koopman, J., Mai, K., De Cremer, D., Zhang, J. H., Reynders, P., Ng, C. T. S., & Chen, I. (2023). No person is an island: Unpacking the work and after-work consequences of interacting with artificial intelligence. Journal of Applied Psychology, 108(9), 1245–1270.

Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. PNAS Nexus, 3(9), pgae346. https://doi.org/10.1093/pnasnexus/pgae346

Tennant, J., Bauin, S., James, S., & Kant, J. (2018). The evolving preprint landscape: Introductory report for the Knowledge Exchange working group on preprints.

Tenzer, H., Abidi, O., & Feuerriegel, S. (2025). Designing LLMs for cultural sensitivity: Evidence from English-Japanese translation. arXiv preprint arXiv:2509.11921.

Tong, J., Sun, Y., Hubbard, R. A., Saine, M. E., Xu, H., Zuo, X., … & Chen, Y. (2025). Incorporating preprints in systematic reviews: a preliminary study of a novel method for rapid evidence synthesis. Journal of the American Medical Informatics Association, 32(11), 1654-1663.

UNESCO, T. (2021). UNESCO recommendation on open science. United Nations Educational, Scientific and Cultural Organization.

Urbaite, G. (2025). AI-Mediated English: How Generative Systems Reinforce English as a Global Lingua Franca. Porta Universorum, 1(9), 35-50.

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. Trends in cognitive sciences, 22(3), 213-224.

Van Bavel, J. J., Rathje, S., Vlasceanu, M., & Pretus, C. (2024). Updating the identity-based model of belief: From false belief to the spread of misinformation. Current Opinion in Psychology, 56, 101787.

Vasista, I., Mirza, I., Huang, C., Patil, R. R., Akalin, A., Zhu, K., & O'Brien, S. MALIBU Benchmark: Multi-Agent LLM Implicit Bias Uncovered. In ICLR 2025 Workshop on Building Trust in Language Models and Applications.

Wang, C. (2024). Exploring students' generative AI-assisted writing processes: Perceptions and experiences from native and nonnative English speakers. Technology, Knowledge and Learning, 1-22.

Wang, J. Y., Sukiennik, N., Li, T., Su, W., Hao, Q., Xu, J., … & Li, Y. (2024). A survey on human-centric llms. arXiv preprint arXiv:2411.14491.

Wang, K., Lyu, T., Su, G., Geiping, J., Yin, L., Canini, M., & Liu, S. (2025). When Fewer Layers Break More Chains: Layer Pruning Harms Test-Time Scaling in LLMs. arXiv preprint arXiv:2510.22228.

Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., … & Wen, Q. (2024). Large language models for education: A survey and outlook. arXiv preprint arXiv:2403.18105.

Wong, P. H. (2025). Global Governance of AI, Cultural Values, and Human Rights. A Companion to Applied Philosophy of AI, 359-371.

Wu, A., Kuang, K., Zhu, M., Wang, Y., Zheng, Y., Han, K., … & Zhang, K. (2024). Causality for large language models. arXiv preprint arXiv:2410.15319.

Yang, R., Tong, J., Wang, H., Huang, H., Hu, Z., Li, P., … & Hong, C. (2025). Enabling inclusive systematic reviews: incorporating preprint articles with large language model-driven evaluations. Journal of the American Medical Informatics Association, 32(11), 1718-1725.

Yazar, B. K., Şahın, D. Ö., & Kiliç, E. (2023). Low-resource neural machine translation: A systematic literature review. IEEE Access, 11, 131775-131813.

Ye, X. (2025). Digital technologies and identity negotiation: a study of trilingual Uyghur university students' language learning experiences in intranational migrations. ReCALL, 37(2), 232-249.

Zangana, H. M., Amelia, P., Mustafa, F. M., & Li, S. (2025). The Role of English Language and AI in Scientific Writing: Ethical and Academic Implications. In Ensuring Secure and Ethical STM Research in the AI Era (pp. 191-218). IGI Global Scientific Publishing.

Zeng, J., & Yang, J. (2024). English language hegemony: retrospect and prospect. Humanities and Social Sciences Communications, 11(1), 1-9.

Zeng, W., Zhu, H., Qin, C., Wu, H., Cheng, Y., Zhang, S., … & Xiong, H. (2025). Application-Driven Value Alignment in Agentic AI Systems: Survey and Perspectives. arXiv preprint arXiv:2506.09656.

Zhang, W., Kam-Kwai, W., Xu, B., Ren, Y., Li, Y., Feng, Y., … & Chen, W. (2025, October). Cultiverse: Towards cross-cultural understanding for paintings with large language model. In Proceedings of the 33rd ACM International Conference on Multimedia (pp. 6710-6719).

Zheng, W. (2024). AI vs. Human: A Comparative Study of Cohesion and Coherence in Academic Texts between Human-Written and ChatGPT-Generated Texts.

Zohouri, M., Sabzali, M., & Golmohammadi, A. (2024). Ethical considerations of ChatGPT-assisted article writing. Synesis (ISSN 1984-6754), 16(1), 94-113.

Lijing Gao is an Assistant Professor of Agricultural Science Communication at the University of Missouri, USA. Her research examines science and risk communication, with a focus on how language, cultural contexts, and social values shape public meaning-making, attitudes, and behavioral responses to emerging technologies. Her current work explores public perceptions and discourse surrounding controversial technology. She has published in leading journals, including International Journal of Communication, Agriculture and Human Values, Precision Agriculture, and Rural Sociology.

Ruanjia Liu is an Assistant Professor of Practice in Accounting at Moravian University, US. Her research focuses on governmental accounting, not-for-profit accounting, and accounting education. Her recent work focuses on using natural language processing to analyze the 10-K filings of U.S. firms. Her job market paper was published in the peer-reviewed Journal of Information Systems (JIS).

Journal of Language, Media and Society
Vol. 1, No. 1 https://doi.org/10.66056/jlms.2026.v1i1.003

Carter & Co. PUBLISHING

RUMOR AS CRISIS DISCOURSE: MEANING-MAKING AND MICRO-RESISTANCE IN SHANGHAI'S DIGITAL PUBLIC SPHERE

YU XIANG

This article examines how digital rumors functioned as crisis discourse during the 2022 Shanghai lockdown, serving both as improvised meaning-making and as fragmented acts of micro-resistance. Drawing on digital ethnography and discourse analysis, the study investigates how residents in a middle-class compound used WeChat groups to circulate, interpret, and act upon rumors amidst strict state censorship and material deprivation. Rather than approaching rumors as mere misinformation, the article conceptualizes them as emergent discursive practices that filled communicative voids, generated grassroots knowledge, and temporarily disrupted dominant state narratives. Grounded in Gramsci's notion of *common sense* and Shibutani's theory of improvised news, the analysis highlights the dialectical nature of rumor as both a survival mechanism and a contested form of bottom-up discourse in authoritarian settings. Framed within the global condition of the post-truth era, this study foregrounds the role of digital platforms specifically WeChat as sites where discourse, power, and control are simultaneously produced, circulated, and contested. In contexts where traditional information infrastructures are compromised, platforms become critical battlegrounds for meaning-making, where rumor emerges as a form of user-generated epistemology. The Shanghai case offers broader insights into how platform architectures, algorithmic visibility, and moderation regimes shape the formation and suppression of alternative discourses during crises. By tracing the micro-politics of rumor in Shanghai's digital public sphere, this article contributes to transnational debates on crisis communication, platform governance, and the shifting dynamics of voice and resistance in digitally mediated authoritarian and post-authoritarian societies.

KEYWORDS: digital rumor; crisis communication; micro-resistance; WeChat; Shanghai lockdown; digital public sphere; authoritarianism

Introduction

Despite a growing body of literature on rumor, grassroots resistance, and digital authoritarianism, few studies have examined how rumor operates as a discursive form of agency within platformized authoritarian infrastructures. While existing work has established the social and emotional functions of rumor during crises (Rosnow, 1988; DiFonzo & Bordia, 2007), less attention has been paid to how rumors not only respond to uncertainty, but actively reconfigure platform communication as a site of hegemonic struggle. Moreover, the interplay between vernacular rumor practices and platform-level governance mechanisms remains under-theorized. This study contributes to these gaps by theorizing rumor as a form

© 2026 Carter & Co.

of discursive improvisation and micro-resistance in the context of China's 2022 Shanghai lockdown. By tracing the circulation of informal messages across WeChat groups, it shows how users simultaneously reproduce and subvert dominant crisis narratives. This approach integrates theories of crisis discourse, common sense, and micro-resistance to offer a grounded account of how meaning and power are negotiated in the digital public sphere under authoritarian conditions.

Background: Zero-Covid Policy, Shanghai Lockdown and Wechat

China's Zero-Covid policy, initiated in early 2020, was characterized by extensive testing, centralized quarantine, and large-scale lockdowns. Unlike many other countries that gradually shifted toward coexistence with the virus, China maintained a strict elimination strategy until late 2022. This approach involved top-down crisis management and high levels of state intervention in everyday life. The 76-day lockdown of Wuhan in 2020 was the first full-city shutdown in modern history (Corradetti & Pollicino, 2021), and it set a precedent for subsequent policies across the country. The governance model underpinning this policy relied heavily on information control, legal enforcement, and the mobilization of digital infrastructures to ensure compliance. Official media channels promoted narratives of scientific necessity and patriotic sacrifice, while alternative discourses, including those questioning quarantine conditions, supply chains, or government competence, were frequently censored (Mendis & Wang, 2020). Public hospitals were converted into quarantine camps, and police-enforced isolation measures were extended to entire residential buildings. During these lockdowns, access to accurate, timely, and transparent information was systematically restricted, creating a communicative void in which rumor and speculation flourished (Zhao & Xiang, 2022; Zhang et al., 2020).

The 2022 lockdown of Shanghai represented a significant rupture in China's pandemic governance. As the country's most cosmopolitan and economically advanced city, Shanghai had previously adopted a more measured strategy emphasizing "precise control" rather than sweeping closures. Yet, when Omicron cases surged in late March, the city was subjected to an indefinite, citywide lockdown without formal announcement or transparent policy communication, generating widespread anxiety, anger, and confusion (Wang & Niu, 2022; Zheng, 2022). Material shortages quickly followed: residents were unable to leave their homes, delivery systems collapsed, and government relief packages were erratic. At the same time, platform-based communication became the only viable method of coordination and survival. In the absence of reliable official messaging, residents turned to online rumors—circulating through WeChat, Douyin, and Weibo—to interpret policy shifts, coordinate food access, and share information about quarantine conditions. In this communicative vacuum, rumor became a form of both epistemic improvisation and everyday resistance.

WeChat, China's dominant all-in-one communication platform, played a central role in both enabling and constraining discourse during the lockdown. As previous studies have shown, WeChat's affordances such as group chats, embedded mini-programs, and synchronized payment systems allowed residents to organize bulk purchases, seek medical help, and track infections informally (Qian & Hanser, 2021). These features made WeChat indispensable for daily survival under lockdown conditions. Yet WeChat also functions as an arm of the state's surveillance apparatus. Group chats are monitored; administrators (group owners) can be held legally accountable for the content shared by members (Zhang, 2018). Accounts that disseminated "sensitive" information were frequently suspended or deleted. This dual role, as a site of grassroots coordination and digital authoritarian control, made WeChat a battleground of discursive tension. Rumors circulating within this hybrid space were not merely unverified claims, but discursive responses to state opacity, reflecting public attempts to reclaim voice, visibility, and meaning in the face of institutional silence.

Theoretical Framework: Rumor in Crisis, Common Sense and Resistance

In crisis situations where official information is delayed, withheld, or distrusted, individuals often turn to informal communicative practices to make sense of uncertainty. Rumors, in such contexts, are not merely distortions or noise in the information environment; rather, they serve as *improvised discourse*—a means through which communities negotiate ambiguity and construct provisional knowledge in the absence of trusted authorities. As Shibutani (1966) famously theorized, rumors are "a recurrent form of communication through which people caught in ambiguous situations attempt to construct a meaningful interpretation by pooling their intellectual resources" (p. 17). He conceptualizes rumor not as an anomaly, but as a *substitute news system*, emerging in moments of institutional breakdown. During the Covid-19 pandemic, this interpretive function of rumor became especially visible in authoritarian settings such as China, where information was tightly controlled. As Zhao and Xiang (2022) observe, health-related rumors circulated widely in digital spaces due to public anxiety, unclear messaging, and fragmented authority. Similarly, Zhang et al. (2020) demonstrate how rumors during the outbreak filled crucial informational voids, enabling individuals to take action in response to perceived threats. These studies support Rosnow's (1988) contention that rumors are best understood as "social constructions" that serve to reduce anxiety and promote group cohesion in high-stress contexts. Rumors in times of crisis thus perform a meaning-making function. As DiFonzo and Bordia (2007) explain, individuals engage with rumors not only to transmit information, but to *interpret and reframe experience*, often collaboratively. In this sense, rumors resemble what Wetherell (1998) describes as "discursive practices that both reflect and shape social realities," particularly under volatile or unstable conditions. While mainstream public discourse may attempt to suppress rumor as irrational or subversive, the persistence of rumor in crisis contexts points to a deeper epistemic need: the creation of shared sense in the face of official silence or contradiction.

While rumor offers an improvised form of meaning-making during crises, it also operates within broader ideological terrains. In Gramsci's framework, "common sense" (*senso comune*) refers to the diffuse, taken-for-granted beliefs and practical knowledges that organize everyday understanding of the world (Gramsci, 1971: 323–324). Though often fragmented and contradictory, common sense is not politically neutral; rather, it is a key terrain on which ideological hegemony is reproduced. Through institutions such as the media, education, and law, dominant groups embed their worldview into what appears "natural" or "self-evident" (Gramsci, 1971: 333–334). In authoritarian settings like China, this process is especially visible during crises, when official discourse seeks to maintain narrative dominance by framing obedience, sacrifice, and state-led control as both rational and patriotic. Yet as Gramsci noted, moments of crisis can rupture hegemonic stability, opening space for alternative forms of common sense to emerge (Gramsci, 1971: 210). When dominant narratives lose credibility or fail to provide answers, subaltern communities may generate competing explanations to

interpret their reality—what Gramsci would call "organic" or counter-hegemonic sense-making (ibid.).

Rumors in this context become discursive instruments for articulating everyday critique. They serve not only to fill epistemic gaps, but also to challenge the coherence and legitimacy of state discourse. As Liu (2016) argues in his study of environmental protest and rumor in China, grassroots communities often develop "vernacular discourses" that question official claims without directly confronting the ideological order. These informal accounts reflect subaltern knowledge practices that operate beneath the threshold of open dissent, yet signal collective skepticism and resistance (Liu, 2016: 22–27). Similarly, O'Brien and Li (2005) describe such engagements as forms of "rightful resistance"—oppositional practices that invoke the state's own rhetoric and rules to critique its failures, thereby creating ideologically compliant, yet politically charged, counter-discourses (O'Brien and Li, 2005: 238–239). In authoritarian societies where overt protest is constrained, resistance often manifests through subtle, everyday practices embedded within routine communication. James Scott's (1985) concept of *"everyday forms of resistance"* provides a foundational framework for understanding such acts not as organized movements but as "infra-political" gestures that operate below the radar of formal power. These include evasions, code-switching, ambiguity, and rumor circulation. They are often uncoordinated, fragmentary, and temporally limited, yet they allow marginalized actors to express dissent and negotiate state authority within the bounds of control (Scott, 1985: xvi–xvii). In the context of China's digital media landscape, such micro-resistance often unfolds within platformized environments that simultaneously enable and regulate communication. Platforms like WeChat, which function as hybrid spaces for commerce, communication, and governance, provide affordances for peer-to-peer interaction but also incorporate extensive surveillance, censorship, and content moderation mechanisms (Han, 2018; Qiang, 2019). These systems limit explicit protest while allowing for ambiguous and indirect modes of oppositional discourse, such as the circulation of emotionally charged rumors or irony-laden posts.

In sum, rumors in crisis contexts are not merely aberrations of rational communication, but culturally and politically situated discursive acts. They function simultaneously as epistemic improvisations and as vehicles of subtle resistance, embedded in everyday vernacular and digital repertoires. In authoritarian environments like China, such discourses allow ordinary citizens to both make sense of volatile realities and negotiate the constraints of hegemonic narratives. Recognizing rumors as sites of meaning-making and micro-contestation invites a methodological approach attuned to the textures of language, context, and mediated interaction. It is with this orientation that the present study undertakes a discourse-centered analysis of rumor practices during the Shanghai lockdown, grounded in digital ethnography and critical interpretive frameworks.

Methodology and Data

This study adopts a digital ethnographic approach to examine rumor circulation and discursive resistance during the Shanghai lockdown in 2022. The primary field site is a middle-class residential compound in Dachang Town, Baoshan District, located on the western edge of Shanghai. Established in 2018, the compound comprises 1,075 households across 48 buildings and is representative of the new suburban middle-class communities in Chinese megacities. The site was chosen for its hybrid characteristics: its residents are economically

stable and digitally literate, yet faced severe constraints during the lockdown, including food shortages, movement restrictions, and limited access to reliable information. By focusing on this site, the research captures how digital rumors function as tools of survival and resistance in an authoritarian crisis context. The study foregrounds the tension between trust and control, where residents are caught between state narratives and everyday uncertainties, and where meaning-making unfolds within highly mediated environments.

The data corpus consists of interactions collected from twelve WeChat groups, as listed in Table 1, active during the lockdown period (April to June 2022). These groups served various functions, from organizing bulk purchases of food and medical supplies to sharing policy updates, neighborhood gossip, and emotional support. The groups ranged in size from under 100 to over 400 members, and daily message volumes varied from approximately 200 to over 1,000. Participant observation was conducted in real-time during the lockdown and continued retrospectively through saved chat histories. Observations included textual messages, emojis, images, voice notes, and forwarded screenshots or external links. Offline behaviors were also noted, particularly how group discussions translated into real-world actions such as coordinated bulk purchasing, rumor-based resource stockpiling, or self-imposed quarantine measures.

The key rumor categories identified through initial inductive coding indicated in Table 2 which are: 1. Food and supply shortages; 2. Omicron outbreaks; 3. Infections within the compound; 4. Covid testing protocols; 5. Quarantine policies; 6. Timeline for lifting restrictions. This typology reflects both the immediate concerns of survival and the broader socio-political uncertainties shaping everyday life. All observations were conducted with attention to ethical research practices. Informed consent was obtained from key members in each WeChat group, and participants were provided with a clear explanation of the research purpose. Consent forms were distributed and signed digitally. To preserve privacy, all identifying information was anonymized, and the researchers refrained from any form of participation that could influence group dynamics.

Table 1. Twelve Internal WeChat Groups of the Compound

Code	Name of Chatgroups	Number of Users	Average Number of Daily Messages (Text, Images, Videos and Others)
SHM	"Second-Hand Market"	435	807
GBFL	"Group-buying: Flour"	267	586
GBMM	"Group-buying: Meat and Milk"	261	653
NB	"Neighborhood"	254	1089
NGB	"Neighbor Group-buying"	210	866
CPN	"Covid Prevention Necessities"	189	230
GBFR	"Group-buying: Fruit"	149	389
TPS	"Toilet Paper &Sanitizer"	123	265
GBR	"Group-buying: Rice"	91	233
GKS	"Group for Kids Stuff"	89	198
No.#	"No. # Houseowners"	87	320
CDH	"Charitable Donation for HOA"	76	228

Given the heightened sensitivity of the context, special care was taken to avoid collecting or reporting any information that could be traced back to specific individuals.

The research adhered to institutional ethical guidelines regarding digital data collection and participant confidentiality. The analysis combined thematic coding with discursive interpretation, focusing on how rumor texts functioned as improvisational responses to uncertainty and tools of everyday resistance. Each rumor was categorized not only by topic, but also by its social function whether it triggered group action, fostered anxiety, strengthened solidarity, or signaled distrust toward state narratives. The coding process tracked rumor trajectories: emergence, amplification, reinforcement (often via individuals with industry contacts), and resolution (clarification or debunking). Attention was also paid to the aesthetics and affect of rumor discourse such as voice notes conveying panic, or humor used to mask critique. The WeChat platform itself was analyzed as a discursive infrastructure both enabling and constraining. It facilitated rapid peer-to-peer rumor circulation but simultaneously hosted embedded surveillance, moderation algorithms, and pro-government narratives. These platform affordances shaped not only what could be said but also how it was said—often in coded, ironic, or ambiguous forms.

Table 2. Six Major Rumor Topics Circulated in WeChat Groups

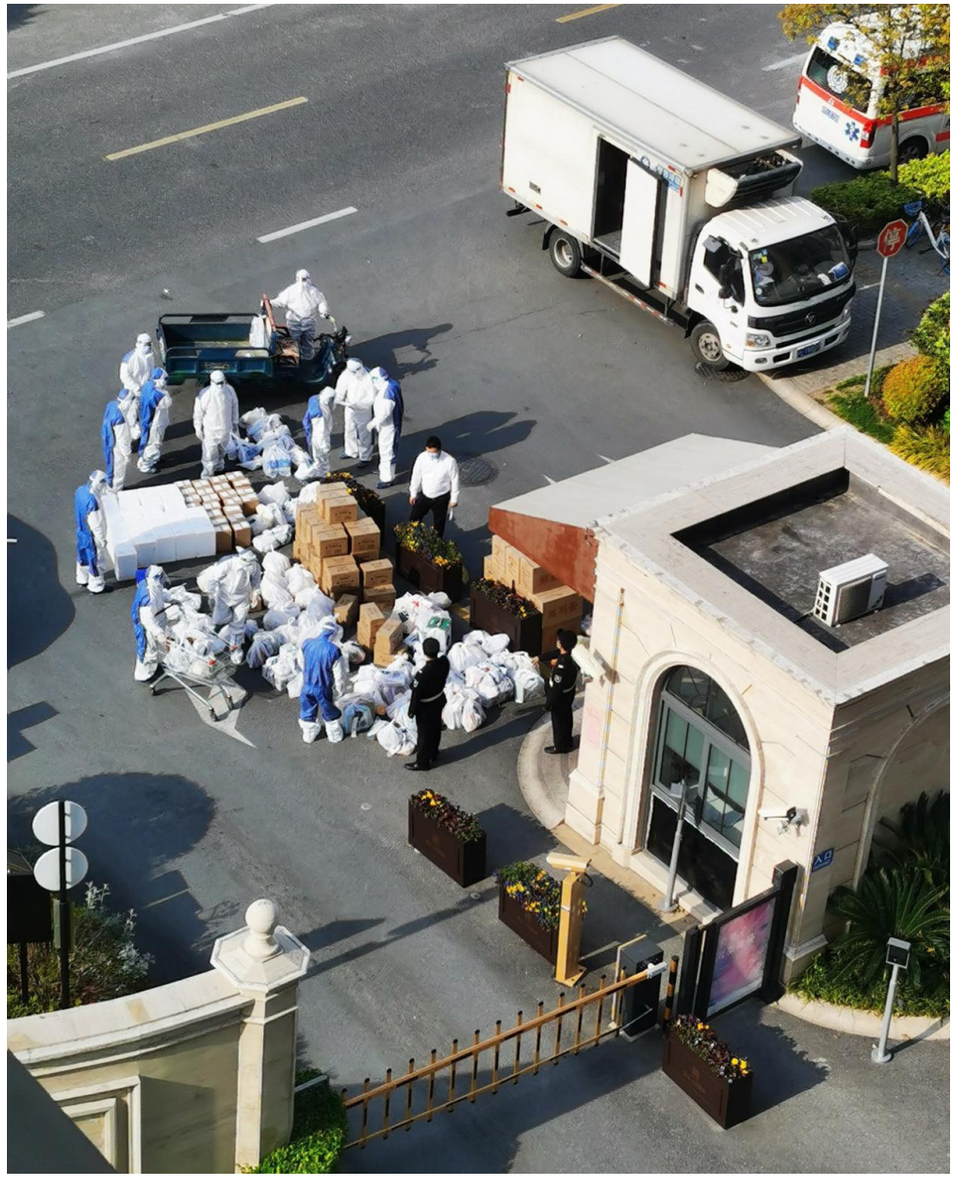
Themes	**Specific Subjects with Examples**			
Food	*Government Supplies Corruption*	*Accessible Platforms to Get Food*	*Fruits and Fast-Food Orders Carrying Virus*	*Necessity of Stocking up Food*
	Town mayor sold governmental supplies to residents in other districts	Tricks to order online groceries; Shanghai government food aid website	Strawberries and KFC are contaminated by deliverymen to spread Covid virus	Government shuts down roads/bridges to suburban farms
Omicron	*Transmission*	*Symptoms & Side Effects*	*Treatment*	*Lethality*
	Long-distance and persistent dispersion through the air	Loss of taste and smell; Pneumonia; Paralysis;	Traditional Chinese medicine	Invalid Chinese vaccines
Infection in Compound	*Infected Buildings*	*Infected Locations*	*Returned Infected Residents*	
	Buildings with infected residents	The units and Apt numbers of infected	Refusal of returned infected from quarantine camps	
Covid Testing	*Medical Staff*	*Testing Method*	*Results*	
	Non-Shanghai medical staff are rough and rude	Different accuracy of single tube and mixed tube sampling	Fake/ Wrong results	
Quarantine	*Quarantine Policy*	*Living Condition*	*Isolation Time*	
	Separate quarantine without parents for children under seven; Compulsive in-house sanitization; Extermination of pets (infected or not).	Unfinished construction buildings without bed and any appliances; Temporary cabin with cracked roof and broken toilets	Minimum 14 days if test negative after admission	
Lifting Date	*Cues from Official News*	*Other Compounds and Districts*	*Speculation and Hearsays*	
	Military troops from other provinces to suppress protests	People in eastern Shanghai went on streets to buy food	The Lockdown continues to October until Xi gets re-elected for third term	

This research is limited in several respects. First, the focus on a single, middle-class compound means the findings may not generalize to other demographics, such as working-class or rural communities. Different socio-economic contexts may produce different rumor ecologies and discursive strategies. Second, the reliance on digital data, while rich in interactional detail, may overlook non-digitally mediated forms of resistance or information flow. Lastly, the Shanghai context, with its high digital penetration and relatively vocal middle class, may differ significantly from other Chinese cities with varying governance and media ecologies.

Rumor as a Survival Infrastructure

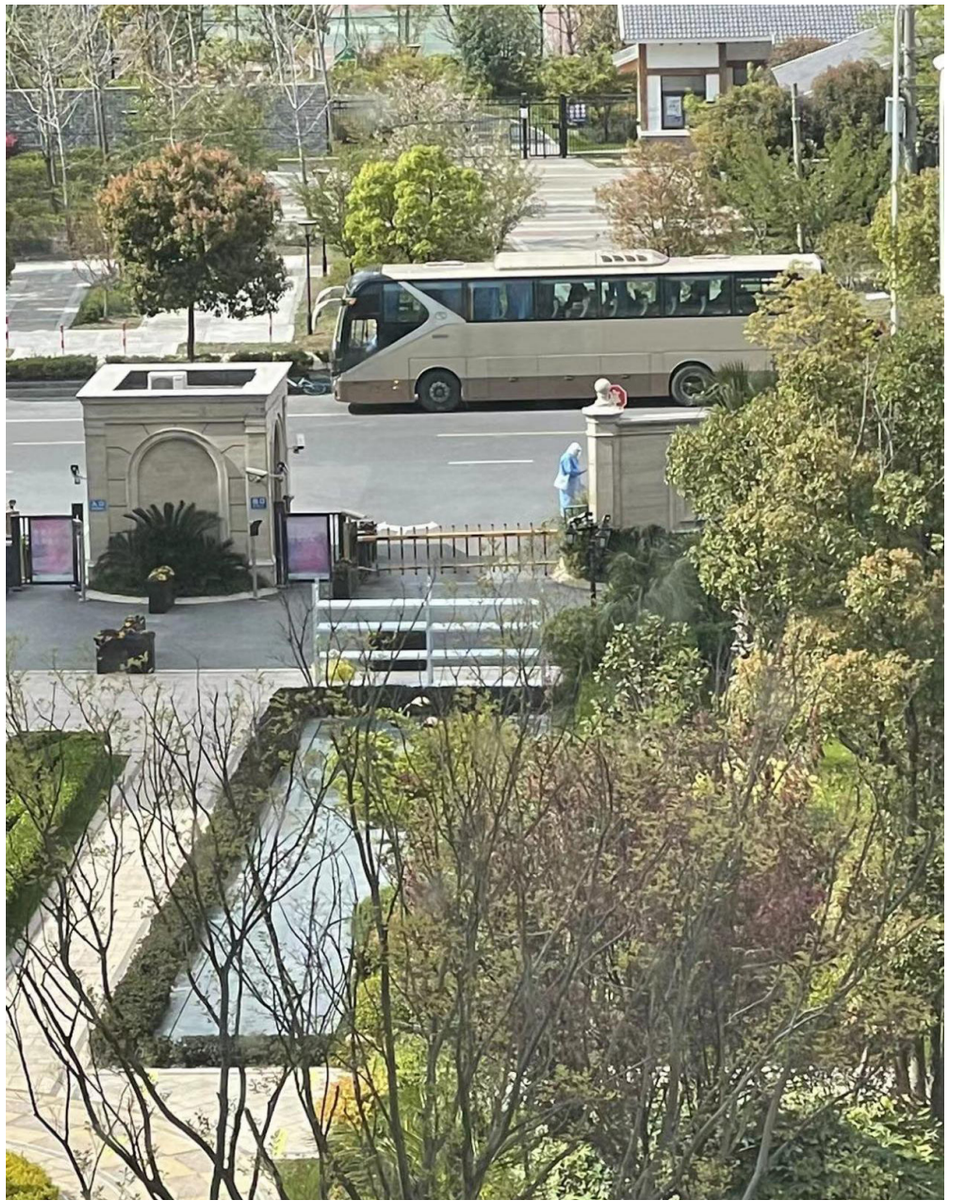
In the first week of the Shanghai lockdown, residents inside the compound repeatedly refreshed the government delivery app, waiting for the promised relief packages. On April 2, one resident shared a photo in the WeChat group "Neighborhood" showing a food box labeled for "Dachang Town" found in Pudong. Several users immediately speculated that these packages had been rerouted. One resident responded: "Maybe the Dachang mayor sold our supplies to his relatives in Pudong." While the rumor could not be verified, it triggered a practical reaction. In a message posted at 7:38 p.m., a group admin in "Group-buying: Flour" announced, "From now on, let's stop waiting. I have a bakery friend with extra inventory—we can organize a bulk order." Within 24 hours, the group grew by over 100 members, each volunteering to coordinate transport, payments, and packaging. Messages were pinned with order forms and digital receipts.

Rumor here was not merely misinformation. It functioned as a mobilizing narrative that gave residents interpretive clarity and a rationale for switching to informal procurement. Rather than awaiting an opaque state response, residents began to build self-organized logistics through preexisting social ties, often using industry connections and alumni networks. This behavior demonstrates how digital rumor operated as a bottom-up meaning-making device in times of uncertainty. It did not require verification to be effective; its real function lay in coordinating material responses under conditions of deprivation. By April 6, multiple WeChat groups were running parallel procurement efforts. Residents shared success stories of food delivery in chat logs, offering validation and encouragement to others. In one post, a resident wrote, "Got my milk today. Thank goodness for group-buying. If we waited for the government, my kid would starve."



Pic 1. Group Purchase – Photography by Anonymous Resident

On April 12, a resident posted a short video in the "Covid Prevention Necessities" WeChat group. The clip, taken from Douyin, showed three toddlers crowded onto a single hospital bed in what appeared to be a quarantine facility in Jinshan District. Their faces were blurred, but the crying was unmistakable. The post sparked a flurry of reactions. Although no death was confirmed, this unverified rumor spread rapidly through several groups. A mother in the "Group for Kids Stuff" group wrote, "Our whole family is locked down. I just pray my child doesn't test positive—we can't let him be sent to a place like that."



Pic 2. Health Control Bus for Covid Positive Resident – Photography by Anonymous Resident

In these exchanges, rumor acted as an emotional intensifier—a means through which individual fear became collective panic. Parents began discussing how to avoid mass testing, fearing that a single positive result would lead to forced family separation. In "Neighborhood," one father posted, "we can't be sent to quarantine." Others agreed, privately messaging their plans to hide symptoms and avoid government reporting.

Pic 3. Video shared in WeChat Groups Where were Believed to be Temporary Quarantine Locations for Covid Positive Patients

The health rumor ecology extended beyond children. Images circulated in multiple groups showing quarantine sites in disrepair: unfinished buildings with no beds, toilets leaking across the floor, and communal spaces without basic ventilation. One user claimed, "That place was just a newly built construction site, not a hospital at all," one user claimed." This was accompanied by a photo of a half-plastered wall, supposedly from the site. These unverified rumors produced actionable behavior. Residents began to reframe the threat of infection: not as a health issue, but as a logistical trap. Infection meant detention. Safety became not a matter of hygiene, but of avoiding the attention of health authorities. In effect, public health discourse had inverted: people feared the cure more than the illness. In this context, rumors functioned not merely as falsehoods but as situated, affective knowledge expressing what official channels refused to acknowledge and triggering behavior consistent with a survival-first mindset. The trauma embedded in the rumor stories (especially those involving children) sharpened community-wide efforts to avoid detection and minimize contact with state-managed health systems.

In addition to concerns about quarantine sites, the fear of infection itself was exacerbated by rumors that visualized and spatialized viral spread. A notable example is the case of "Strawberry Cluster" at Shangnan Third Village, where an image widely circulated on WeChat and other messaging platforms claimed that the entire neighborhood had been infected due to a COVID-positive fruit vendor. The post included a detailed residential map annotated with infected zones, highlighting the supposed epicenter and drawing an implied causal link between the infected vendor and community-wide transmission. Such rumors were often accompanied by emojis (e.g., crying, facepalm) that simultaneously expressed humor

and despair, suggesting a blend of helplessness and sarcasm. These user-generated "infection maps," although not officially verified, served as an informal epistemology of risk, guiding residents' behavior and reinforcing hyper-vigilance. By offering a tangible visualization of threat, rumors like this transformed abstract anxieties into actionable boundaries. Residents would avoid certain blocks, refuse deliveries from specific areas, or even blame individuals for community spread. These discursive artifacts did not simply misinform. They reorganized the mental geography of risk during lockdown.

Pic 4. Screenshot of Covid Positive Strawberries

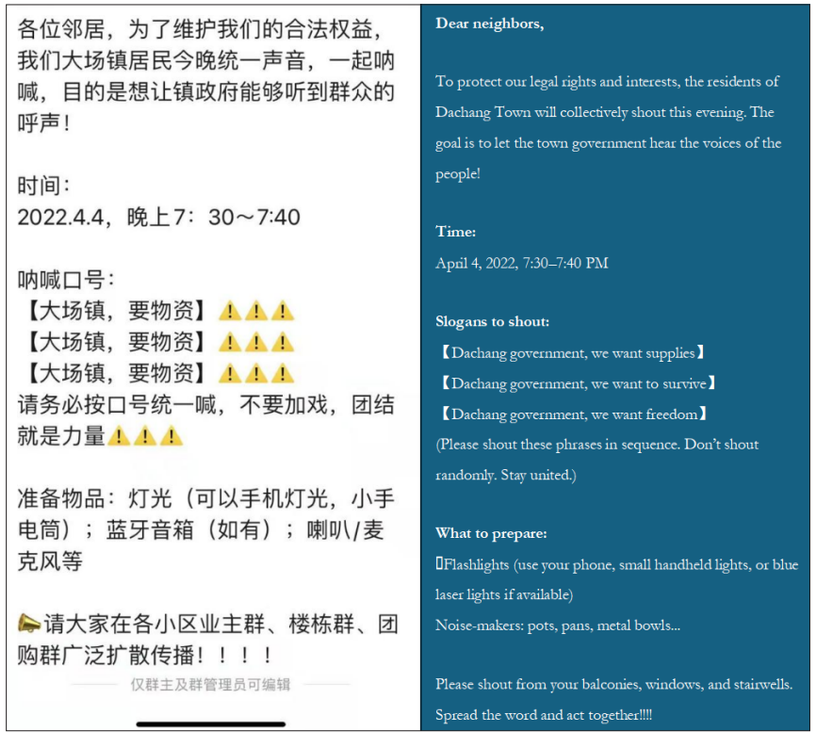
Rumor as Disputed Resistance Tool and Limitations

As the lockdown persisted, digital rumors also began to serve as tools for resistance against governmental control. The Baoshan Corruption incident did not only urge the residents to form their own food supply infrastructure but also pushed some online protest movements in the district. Screenshots of Weibo posts under the hashtag "Baoshan Corruption" began to appear. The hashtag, at its peak, was reportedly viewed over 5 million times in a single day, although many posts were swiftly taken down. A resident of Jingbei Compound shared a screen recording of them trying to repost the hashtag, only to be met with an error message

and then a temporary account suspension. Beyond online resistance, rumors also played a role in coordinating offline protests. For example, when residents in a nearby Gu Village organized a protest, the details were shared through WeChat, including textual descriptions and pictures of the event. This sparked a sense of solidarity among other compounds, and soon after, a video of people shouting in the streets, assumed to be from Dachang, went viral within the compound. Though it was later discovered that the protest had actually taken place in Qingdao, a northern city, the rumor nonetheless inspired local residents to plan similar collective protests.

Pic 5. Video shared in WeChat Groups of Protest in another town named Da Chang

One notable offline action was a proposed "Han Lou" (喊楼) – collective shouting campaign, where residents were encouraged via WeChat to shout slogans from their windows at a designated time. However, the plan caused immediate concern within the group, with residents posting comments such as, "Let's quickly leave more messages to roll that up" and "Don't send such dangerous information anymore." The panic was fueled by a 2017 law that held group advocates legally responsible for their members' actions, resulting in hesitation to engage in any potentially illegal activities (Zhang, 2018). This led to a culture of self-policing within WeChat groups, stifling dissenting voices before they could escalate into substantial movements. As the government tightened control, many WeChat accounts and groups that shared dissident views were either banned or "blown up" (炸号). This was a devastating blow to residents, as WeChat was their primary platform for accessing up-to-date information and coordinating survival efforts. While the Baoshan corruption protests led to small victories, such as the eventual delivery of food supplies, the larger-scale resistance against the zero-Covid policy was fragmented and ultimately unsuccessful. The use of rumors as a resistance tool became increasingly limited as the government's censorship mechanisms and WeChat surveillance curbed the spread of dissenting information, leading to a short-lived and disjointed resistance movement. The call for collective shouting from balconies ultimately transformed, within the sampled compound in this study, into a collective balcony performance of the patriotic song "I Love You, China."



Pic 6. Screenshot shared in WeChat Groups about "Han Lou"

The tension between resistance and compliance is constantly shifting and mutating. Another example is the case of nucleic acid testing which exemplifies how even small acts of resistance were quickly suppressed. When residents in one building refused to go downstairs for testing after hearing rumors that someone in their building was infected, they demanded that the medical staff conduct tests at their doors instead. This minor standoff, however, was short-lived as residents feared the consequences of defying government orders, particularly losing access to their "health codes" (健康码), which were required for daily life. The fear of legal repercussions and government sanctions effectively curtailed even minor forms of resistance. In the final weeks of lockdown, the atmosphere in WeChat groups shifted from open resistance to compliance, as residents prioritized survival over protest. With food channels stabilized and residents gaining access to selective luxury items like tiger prawn, civet coffee, and La Mer, the once-strong voices of dissent faded into silence. However, small-scale protests persisted in specific contexts, such as when residents of other districts smashed rotten food in the streets or clashed with police over forced home expropriations for quarantine sites. In these cases, the rumors spread on WeChat were more about specific grievances rather than calls for broader resistance, further highlighting the fragmented nature of digital protest under the surveillance state.

Constructing Informal Truths: Digital Rumor as Everyday Meaning-Making

The following picture is one of the clearest indications of the Shanghai government's loss of public trust during the lockdown emerged not from what was disclosed, but from what

was denied. On March 22, 2022, online speculation about an imminent citywide lockdown—spurred by inconsistent policy signals and rising case numbers—spread rapidly across digital platforms. Rather than providing transparent clarification, the Shanghai Public Security Bureau issued an official bulletin on March 23 stating that a 42-year-old man surnamed Sun had been administratively detained for "fabricating" the claim that "Shanghai will be locked down for 7 days." The statement emphasized the illegality of spreading rumors and warned the public not to "believe or spread false information." Yet only a few days later, the very scenario described in the alleged "rumor" unfolded in full: the city entered a protracted, two-phase lockdown starting March 28. The punitive treatment of preemptive information as "falsehood" laid bare the state's strategy of managing public emotion through suppression rather than clarity. This moment marked a turning point in the credibility of official communication—when citizens realized that not only were rumors sometimes more accurate than government channels, but that truth itself could be retroactively criminalized.

Pic 7. Governmental Announcement of Debunking Fake Rumor of Citywide Lockdown

In authoritarian digital environments, when formal channels of communication are censored or distrusted, rumors become vernacular epistemologies—informal, often fragmentary truth claims that circulate in digitally mediated spaces such as WeChat or Weibo (Liu, 2016). These rumors help ordinary users navigate everyday survival, but also allow for the construction of *alternative knowledge systems*, however precarious. In the context of the Shanghai lockdown, rumors about food availability, quarantine policy, and infection rates did not merely fill gaps left by the state; they became discursive acts of interpretive agency. They enabled individuals to participate in bottom-up knowledge production, thereby momentarily reconfiguring power over meaning.

In digitally mediated authoritarian contexts, such as the Shanghai lockdown, these

grassroots discourses are shaped by and embedded within platform architectures. Platforms like WeChat become arenas of discursive negotiation, where rumors emerge as micro-level textual disruptions of official hegemony. Residents engaging in rumor circulation were not simply reacting to uncertainty, but actively reframing dominant narratives about the Zero-Covid policy, quarantine conditions, and state responsibility. These rumors constituted temporary "counter-publics" (Fraser, 1990: 67) fleeting spaces of oppositional discourse produced within, and constrained by, the dominant communicative regime. Yet this resistance is rarely explicit or sustained. As Gramsci warned, common sense is a contradictory space, often containing both hegemonic and counter-hegemonic elements (Gramsci, 1971: 423). Rumors can destabilize official narratives, but they can also be reabsorbed into dominant discourse. For example, when skepticism mutates into resignation, or when protest devolves into rumor-driven panic. The discursive field is thus dialectical and unstable, shaped by ongoing struggles over meaning, legitimacy, and voice.

O'Brien and Li's (2005) notion of *"rightful resistance"*, therefore, can be extended to the platform-specific forms of resistance observed during the Shanghai lockdown. For instance, residents shared screenshots, hearsay, or videos within semi-private WeChat groups to question the legitimacy of lockdown measures, or to mock state narratives under the guise of concern or confusion. These actions constituted discursive cracks in hegemonic control, made possible by the blurred boundary between private and public communication on Chinese platforms (Herold and Marolt, 2011: 12–14). However, the nature of these digital micro-resistances is inherently precarious. The design of Chinese platforms embeds architectures of responsibilization, where group administrators can be held legally liable for "harmful" content, thus encouraging pre-emptive self-censorship and internal policing (Zhang, 2018). This produces what Yang (2009: 13) calls "controlled inclusion"—a system in which limited expressions of dissent are tolerated, but structurally constrained. As a result, acts of micro-resistance often remain episodic, individualized, and fragile, lacking the continuity or institutional support necessary for long-term transformation.

Nonetheless, these everyday discursive acts such as spreading rumors about food corruption, exaggerating quarantine risks to avoid relocation, or organizing bulk purchases through informal groups can be seen as forms of improvised agency. They reconfigure platform use from passive consumption to strategic survival and discursive contestation, even if only temporarily. The WeChat-based communication during the Shanghai lockdown illustrates how the digital public sphere in China is not a dichotomy of control vs. resistance, but a dialectical space where power and voice are constantly being renegotiated in granular, contingent ways (Dai and Chen, 2022).

Conclusion and Implications

This study has examined how digital rumors operated as both survival tools and improvised counter-discourses during the Shanghai lockdown of 2022. By focusing on everyday communicative practices in a middle-class residential compound, it has highlighted the ways in which residents used rumors to navigate uncertainty, express discontent, and construct localized forms of meaning-making under conditions of heightened control. While many of these acts were ultimately constrained by platform governance and political suppression, the ephemeral and adaptive nature of rumor-based communication points to a deeper cultural

logic of resistance under authoritarian rule. Yet the relevance of this study extends far beyond its original context. Since 2022, global events have continued to demonstrate the centrality of informal, decentralized, and emotionally charged information flows in times of crisis. From the rapid spread of war-related rumors on Telegram during the Ukraine invasion, to AI-generated misinformation during the 2024 U.S. election cycle, to WhatsApp-driven panic during the India-Manipur ethnic violence, we see a transnational pattern of digital rumor becoming an essential feature of contemporary public life. In each case, the blurred boundary between truth, speculation, and affect becomes not an anomaly but a systemic condition of platformed communication.

Theoretically, this reinforces the need to reconceptualize rumor not as a failure of rational discourse, but as a discursive response to epistemic fragmentation, institutional distrust, and emotional urgency. In an age marked by post-truth politics, platform monopolies, and fragmented publics, rumor emerges as an adaptive, situational, and collectively authored mode of knowledge production. It fills the vacuum left by collapsing institutional credibility, offering provisional truths when no stable narrative is available. Rather than dismissing rumor as irrational or dangerous, we must understand it as a vernacular epistemology—a mode of knowing rooted in lived experience, shared affect, and tactical ambiguity. Its fragility and volatility are not signs of its weakness, but of its responsiveness to crisis. As such, the rumor should not be treated merely as an object of regulation or correction, but as a lens through which to understand how publics, especially under authoritarian or uncertain conditions, seek to make sense of the world and to reassert agency, however momentarily, in the face of imposed silence.

References

Corradetti, Claudio, and Oreste Pollicino. 2021. "The 'War' Against Covid-19: State of Exception, State of Siege, or (Constitutional) Emergency Powers?" German Law Journal 22 (6): 1060–1071.

Dai, Jia, and Hao Chen. 2022. "Fragmented Publics: Power, Trust, and Communication in China's Digital Governance." Information, Communication & Society 25 (5): 593–609.

DiFonzo, Nicholas, and Prashant Bordia. 2007. Rumor Psychology: Social and Organizational Approaches. Washington, DC: American Psychological Association.

Fraser, Nancy. 1990. "Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy." Social Text, no. 25/26: 56–80.

Gramsci, Antonio. 1971. Selections from the Prison Notebooks. Edited and translated by Quintin Hoare and Geoffrey Nowell Smith. London: Lawrence and Wishart.

Han, Rongbin. 2018. Contesting Cyberspace in China: Online Expression and Authoritarian Resilience. New York: Columbia University Press.

Herold, David Kurt, and Peter Marolt, eds. 2011. Online Society in China: Creating, Celebrating, and Instrumentalising the Online Carnival. London: Routledge.

Liu, Jun. 2016. "Digital Media, Cycle of Contention, and Sustainability of Environmental Activism: The Case of Anti-PX Protests in China." Mass Communication and Society 20 (5): 604–625. https://doi.org/10.1080/15205436.2016.1203954.

Mendis, Patrick, and Joey Wang. 2020. "The Three Mistakes the Chinese Government Has Made in Its Mishandling of the Coronavirus Crisis." South China Morning Post, February 7, 2020. https://www.scmp.com/comment/opinion/article/3049460/three-mistakes-chinese-government-has-made-its-mishandling.

O'Brien, Kevin J., and Lianjiang Li. 2005. "Popular Contention and Its Impact in Rural China." Comparative Political Studies 38 (3): 235–259.

Qian, Yue, and Amy Hanser. 2021. "How Did Wuhan Residents Cope with a 76-Day Lockdown?" Chinese Sociological Review 53 (1): 55–86.

Qiang, Xiao. 2019. "The Road to Digital Unfreedom: President Xi's Surveillance State." Journal of Democracy 30 (1): 53–67.

Rosnow, Ralph L. 1988. "Rumor as Communication: A Contextualist Approach." Journal of Communication 38 (1): 12–28.

Scott, James C. 1985. Weapons of the Weak: Everyday Forms of Peasant Resistance. New Haven, CT: Yale University Press.

Shibutani, Tamotsu. 1966. Improvised News: A Sociological Study of Rumor. Indianapolis: Bobbs-Merrill.

Wang, Vivian, and Isabelle Niu. 2022. "Shanghai Seethes in Covid Lockdown, Posing Test to China's Leadership." New York Times, April 7, 2022. https://www.nytimes.com/2022/04/07/world/asia/shanghai-covid-lockdown-china.html.

Wetherell, Margaret. 1998. "Positioning and Interpretative Repertoires: Conversation Analysis and Post-Structuralism in Dialogue." Discourse & Society 9 (3): 387–412.

Yang, Guobin. 2009. The Power of the Internet in China: Citizen Activism Online. New York: Columbia University Press.

Zhang, Lifang, Kexin Chen, Haijun Jiang, and Jian Zhao. 2020. "How the Health Rumor Misleads People's Perception in a Public Health Emergency: Lessons from a Purchase Craze during the COVID-19 Outbreak in China." International Journal of Environmental Research and Public Health 17 (19): 7213. https://doi.org/10.3390/ijerph17197213.

Zhang, Di. 2018. "The Group Member Commits a Crime, and the Group Owner Is Implicated." Beijing Daily, October 18, 2018. https://ie.bjd.com.cn/.

Zhao, Xuan, and Yining Xiang. 2022. "Using Disparagement Humor to Deal with Health Misinformation Endorsers: A Case Study of China's Shuanghuanglian Oral Liquid Incident." In The Palgrave Handbook of Media Misinformation, edited by Homero Gil de Zúñiga, Anamaria Neag, and Miriam Steiner, 179–190. Cham: Palgrave Macmillan.

Zheng, Wei. 2022. "The Truth of the Shanghai Model." Yicai, March 16, 2022. https://www.yicai.com/news/101345678.html.

Dr. Yu Xiang is currently teaching at Metro State University in Minnesota, United States. She received her PhD from the University of Westminster in UK. Her current research interest lies in the areas of critical discourse analysis, audience reception and digital culture.

Journal of Language, Media and Society
Vol. 1, No. 1 https://doi.org/10.66056/jlms.2026.v1i1.004

Carter & Co. PUBLISHING

THREE VEHICLES OR FOUR VEHICLES?

A HERMENEUTICAL EXAMINATION OF EARLY INTERPRETATIONS OF THE PARABLE OF THE THREE CARTS

JUN YAN

The Parable of the Three Carts in the *Lotus Sūtra*, also known as the Parable of the Burning House, has been interpreted differently in Chinese Buddhist exegesis, with a division between the Three-Vehicle School (*sanche jia* 三車家) and the Four-Vehicle School (*siche jia* 四車家). The distinction lies in whether the ox-cart among the three carts is identical to the final great white ox-cart, which essentially reflects different understandings of the relationship between the Three Vehicles (*triyāna*) and the One Vehicle (*ekayāna*). This division already existed before the emergence of sectarian Buddhism. Early proponents of the Three-Vehicle interpretation include Huiguan, Sengzhao, Sengrui, Daosheng, and Liu Qiu; those of the Four-Vehicle interpretation include Fayun and Huisi. Fayun represents a crucial turning point, pioneering the Fourth Vehicle interpretation. Huisi, building upon this foundation, used his own contemplative experience and *tathāgatagarbha* theory to develop a second path for the Four-Vehicle School. The fundamental cause for the emergence of the Four-Vehicle School lies in the further polarization of the relationship between expedient means (*upāya*) and reality (*tattva*), which consequently granted the One Vehicle an independent status with concrete content.

KEYWORDS: Parable of the Three Carts; Parable of the Burning House; expedient means and reality (quan-shi 權實); Three Vehicles and One Vehicle

The Parable of the Three Carts appears in the "Parable Chapter" (*Piyu pin* 譬喻品) of the *Lotus Sūtra* (*Saddharmapuṇḍarīka-sūtra*), also known as the Parable of the Burning House. It tells of an elder who sees his children playing in a burning house and lures them out by promising them goat-carts, deer-carts, and ox-carts. After they escape, the elder bestows upon each of them a great white ox-cart. As a tool for explaining doctrine, parables capture only partial resemblances, thus allowing for multiple interpretive possibilities.

The Parable of the Three Carts has two main interpretations: Three Vehicles and Four Vehicles. The distinction lies in whether the ox-cart among the three carts is identical to the final great white ox-cart. Those who hold there are three carts are called the Three-Vehicle School; those who hold there are four are called the Four-Vehicle School. The three carts symbolize the Three Vehicles, while the great white ox-cart symbolizes the One Buddha Vehicle. The fundamental difference between the two schools concerns whether the One Buddha Vehicle is identical to the Great Vehicle (Mahāyāna) among the Three Vehicles.

The Parable of the Three Carts is closely connected to the *Lotus Sūtra*'s central themes of "revealing the real through expedient means" (*kaiquan xianshi* 開權顯實) and "uniting

© 2026 Carter & Co.

the three into one" (*huisan guiyi* 會三歸一). Hence, it has always received special attention from commentators. As Jizang (549-623) noted: "The debate over the three carts has been contentious for a long time. Understanding this enables comprehension of the entire sūtra; confusion about it obstructs all seven scrolls."[1]

Among extant *Lotus Sūtra* commentaries, the most influential are those by Jizang, Zhiyi (538-597), and Kuiji (632-682). Their interpretations had enormous influence in the era of sectarian Buddhism. The Tiantai and Huayan schools belong to the Four-Vehicle School, holding that the One Vehicle is a vehicle beyond the Three Vehicles—for Tiantai, this is the Perfect and Sudden Teaching; for Huayan, it is the Distinct Teaching of the One Vehicle. The Sanlun and Weishi schools belong to the Three-Vehicle School, holding that the One Vehicle is simply the Great Vehicle, with no independent One Vehicle existing separately.[2]

Zhiyi summarized: "People differ in their understanding of the number of carts and the nature of the carts. Some say there are initially three carts, with the later teaching uniting two into one; some say there are initially three, with the later teaching uniting three into one; some say there are initially four, with the later teaching uniting three into one."[3] When the Japanese monk Saichō (767-822) came to Tang China to study, he specifically asked the abbot Daosui of Chanxiu Monastery on Mount Tiantai whether Tiantai belonged to the Three-Vehicle or Four-Vehicle School.[4]

However, the brilliant achievements of Sui-Tang sectarian Buddhism have obscured the fact that the divergence between Three-Vehicle and Four-Vehicle interpretations already existed during the Eastern Jin and Northern and Southern Dynasties, significantly influencing later sectarian exegesis. Therefore, this article focuses on early commentators' interpretations of the Parable of the Three Carts—what might be called the "prehistory" of these interpretations.[5] The different interpretations of this parable contain profound intellectual roots. A hermeneutical examination of early interpretations not only helps deepen our understanding of concepts like expedient means and reality, as well as Mahāyāna thought in Chinese Buddhist intellectual history, but also provides an important entry point for understanding ancient hermeneutics.

The Lotus Sūtra Text as the Interpretive Foundation

No matter how later interpretations develop, they must take the *Lotus Sūtra* text as their foundation. Therefore, we must examine what the sūtra itself says about the Three Vehicles, the three carts, the great white ox-cart, and the One Vehicle. The Three Vehicles and three carts are quite clear in the *Lotus Sūtra*. Kumārajīva's (343-413) translation of the "Parable Chapter" states:[6]

> "Śāriputra! If there are sentient beings who possess inner wisdom, and upon hearing the Dharma from the Buddha World-Honored One, accept it with faith, diligently and earnestly striving to quickly escape the three realms and seeking nirvāṇa for themselves—these are called the Śrāvaka Vehicle, like those children who sought the goat-cart to escape the burning house. If there are sentient beings who, upon hearing the Dharma from the Buddha World-Honored One, accept it with faith, diligently and earnestly striving, seeking natural wisdom, delighting in solitary goodness and tranquility, deeply knowing the causes and conditions of all dharmas—these are called the Pratyekabuddha Vehicle, like those children who sought the deer-cart to escape the burning house. If there are sentient beings who, upon hearing the Dharma from the Buddha World-Honored One, accept it with faith, cultivating with diligent effort, seeking all-knowledge, Buddha-knowledge, natural knowledge, teacherless knowledge, the Tathāgata's seeing and knowing, powers, and fearlessnesses, compassionately thinking of and bringing peace and joy to countless sentient beings, benefiting gods and humans, liberating all—these are called the Great Vehicle. Because bodhisattvas seek this vehicle, they are called mahāsattvas, like those children who sought the ox-cart to escape the burning house."

1 [Sui] Jizang, Fahua xuanlun 法華玄論 [Treatise on the Profound Meaning of the Lotus Sūtra], in Taishō shinshū daizōkyō 大正新修大藏經 (Tokyo: Taishō Issaikyō Kankōkai, 1988), vol. 34, p. 389a.

The Three Vehicles are the Śrāvaka Vehicle, the Pratyekabuddha Vehicle, and the Great Vehicle; the three carts are the goat, deer, and ox carts corresponding to the Three Vehicles. This passage is the foundation for all later interpretations.

In the "Parable Chapter," "the great white ox, fat and strong, of beautiful form, pulling the jeweled cart" corresponds to the One Vehicle. However, the Chinese text allows for a certain interpretive space, which is the origin of different later interpretations. The "Parable Chapter" states: "Just as that elder, seeing all his children safely escape the burning house to a place without fear, reflects on his immeasurable wealth and equally bestows great carts upon all his children... These sentient beings are all my children; I give them equally the Great Vehicle."[7] Here, "equally bestowing great carts" corresponds to "giving them equally the Great Vehicle," meaning that the ox-cart among the three carts is the great white ox-cart, which represents the Great Vehicle.

But then it also says: "Initially using three carts to entice the children, afterward giving only the great cart." And: "Initially teaching the Three Vehicles to guide sentient beings, afterward using only the Great Vehicle to liberate them." And: "He can give all sentient beings the Dharma of the Great Vehicle, but not all can receive it. Śāriputra! For this reason, know that the Buddhas, through the power of expedient means, distinguish and teach three in what is actually the One Buddha Vehicle."[8]

This seems interpretable as meaning that the ox-cart or Great Vehicle among the three carts and Three Vehicles is not the Great Vehicle in the sense of the One Buddha Vehicle. Moreover, in Dharmarakṣa's (231-308) earlier translation, the *Zhengfa hua jing* 正法華經, "giving them equally the Great Vehicle" is rendered as "universally encouraging advancement toward the Buddha Vehicle."[9] The implication that the One Buddha Vehicle is not equivalent to the Great Vehicle seems even clearer here.

Different interpretations of the Parable of the Three Carts focus on the great white ox-cart that symbolizes the One Vehicle; the essential question is how to understand the One Vehicle. The One Vehicle means the One Buddha Vehicle. The "Expedient Means Chapter" says: "The Tathāgata teaches the Dharma to sentient beings by means of only the One Buddha Vehicle; there is no other vehicle, whether second or third." In Sanskrit: *ekam evāhaṃ śāriputra yānamārabhya sattvānām dharmaṃ deśayāmi yad idaṃ buddhayānaṃ | na kiṃci cchāriputra dvitīyaṃ vā tṛtīyaṃ vā yānaṃ saṃvidyate*.[10]

7 [Later Qin] Kumārajīva, trans., Miaofa lianhua jing, T09, p. 13b.

8 [Later Qin] Kumārajīva, trans., Miaofa lianhua jing, T09, p. 13c.

9 [Western Jin] Dharmarakṣa, trans., Zhengfa hua jing 正法華經, T09, p. 76b.

10 H. Kern and Bunyiu Nanjio, ed., Saddharmapuṇḍarīkasūtram (St. Petersburg, 1908), p. 40.

"Whether second or third"—the "whether...or" structure is *vā...vā*; "second" is *dvitīya* and "third" is *tṛtīya*, both ordinal adjectives modifying *yāna* (vehicle) in the neuter singular nominative. Thus, the meaning in Sanskrit is very clear: there is only the One Buddha Vehicle (*ekaṃ buddhayāna*); there is no second or third vehicle. Both Jizang and Kuiji noted this point. However, based on the Chinese wording alone, "two" and "three" can also be interpreted as "two kinds of vehicles" and "three kinds of vehicles."

Extant early Lotus-related literature can be divided into three categories. First, complete *Lotus Sūtra* commentaries—the earliest being the *Lotus Sūtra Commentary* (*Fahua jing shu* 法華經疏) by Zhu Daosheng (355-434), followed by the *Lotus Sūtra Record of Meaning* (*Fahua yiji* 法華義記) by Guangzhai Fayun (467-529). These are the most important documents for understanding early interpretations.[11]

The Early Three-Vehicle School

Early proponents of the Three-Vehicle interpretation include Huiguan, Sengzhao, Sengrui, Daosheng, and Liu Qiu (438-495). Of course, except for Daosheng, the extant documents of these figures are too sparse, leaving some ambiguity; we can only say they are closer to the Three-Vehicle School, not as definitive as Daosheng.

(1) Huiguan and Sengzhao

Huiguan studied under Kumārajīva and was listed as one of the Four Sages of Kumārajīva's school. "In precise debate, Huiguan and Sengzhao ranked first."[12] Huiguan wrote "Preface to the Essential Points of the *Lotus*," which reportedly received Kumārajīva's approval upon completion.[13]

Based on this preface, Huiguan mainly discusses the Three Vehicles and One Vehicle from two levels: the expedient means of teaching and the ultimate reality. The Three Vehicles are "responses to beings that open pathways" and "provisional responses" that "cannot speak of the ultimate to beginners"—they are expedient means for teaching beginners. Hence the Three Vehicles are compared to "separate streams." The separate streams are not reality; three different rivers ultimately merge together, converging into the real One Vehicle of the Wonderful Dharma.

Huiguan does not explicitly state whether there are three or four carts, but from his metaphor—"Ten thousand streams merge and flow together; the Three Vehicles proceed as one. The three that proceed together unite into one"—like three rivers merging into one, where this one river does not exist independently of the three, he is closer to the Three-Vehicle School. Sengzhao, "foremost in understanding emptiness among the Qin," did not write specifically on the *Lotus*, but in his *Treatise on the Namelessness of Nirvāṇa*, he mentions the Parable of the Three Carts:[14] [15]

11 From extant documents, Kumārajīva focused on answering the possibility of arhats becoming Buddha; he did not address the Three Vehicles-One Vehicle relationship or the Parable of the Three Carts.

12 [Liang] Huijiao, Gaoseng zhuan 高僧傳 , T50, p. 368b.

13 [Sui] Jizang, Fahua xuanlun, T34, p. 380a.

14 [Liang] Sengyou, ed., Chu sanzang ji ji 出三藏記集 , T55, p. 57a.

15 [Eastern Jin] Sengzhao, Zhao lun 肇 論 , T45, p. 159c. There is debate about the authorship of the Niepan wuming lun; this article follows the traditional attribution to Sengzhao.

"The Lotus Sūtra says: The first great path has no two destinations. I, through expedient means, for the indolent, distinguish and teach three within the One Vehicle path. The three carts escaping the burning house is precisely this matter. Because all escape birth and death, they are equally called unconditioned; because what they ride differs, there are three names. Unifying their destination, there is only one."

The One Vehicle and Three Vehicles are essentially no different; all can achieve liberation from birth and death. It is only because of different pedagogical arrangements that there are distinctions of Three Vehicles. Therefore, Sengzhao should also be classified with the Three-Vehicle School.

(2) Sengrui

Sengrui was Kumārajīva's most important translation assistant, highly valued by Kumārajīva, and listed among the "Four Sages" and "Eight Outstanding Ones," holding a "leading" position.[16] [17] Sengrui wrote the "Postface to the *Lotus Sūtra*." His distinctive approach was to compare the *Lotus Sūtra* with the *Prajñāpāramitā Sūtras*. He held that although the *Prajñāpāramitā* "reaches the utmost in profundity" and "encompasses everything in its greatness," nevertheless "all take responsive transformation as fundamental... As for transformation through skillful means, although it broadly awakens beings, it is insufficient regarding the true substance—all belong to the *Lotus*."

This shows Sengrui is closer to "other-emptiness" rather than "self-emptiness," not deeply aligned with Mādhyamika thought. However, Sengrui does not elevate the *Lotus* while denigrating the *Prajñāpāramitā*; he harmonizes the two, using the *Prajñāpāramitā*'s thought of emptiness to explain the *Lotus*. Another distinctive feature is that he does not consider the unification of the Three Vehicles into One the most important doctrine. He regards the ideas of the Buddha's attainment of enlightenment in the distant past, the Buddha's immeasurable lifespan, and emanation bodies as most important—already containing the embryonic form of the later "origin teaching" and "trace teaching" division.

(3) Zhu Daosheng

Among extant Chinese *Lotus Sūtra* commentaries, the *Lotus Sūtra Commentary* by Zhu Daosheng, "the Sage of Nirvāṇa," is the earliest. This commentary was completed two years before Daosheng's death, in the ninth year of Yuanjia (432), representing his mature thought.[18]

1. He interprets the Three Vehicles and One Vehicle through expedient means and reality. In explaining the three carts parable, Daosheng says: "The Buddha's transformation operates among humans; the Small Vehicle follows a teacher—these two are compared to ox and goat, which are things of the human realm. Pratyekabuddhas neither transform others nor follow teachers, so they are compared to deer."[19] The Three Vehicles are merely expedient means, provisional rather than real: "Suddenly hearing of the beauty of the three carts, their hearts surely delight in receiving them—this is not real teaching; it is called expedient means."[20]

16 [Sui] Jizang, Zhongguan lun shu 中觀論疏 , T42, p. 1a.

17 [Liang] Sengyou, ed., Chu sanzang ji ji, T55, p. 57b.

18 [Eastern Jin] Zhu Daosheng, Fahua jing shu 法華經疏 , X27, p. 6c.

19 [Eastern Jin] Zhu Daosheng, Fahua jing shu, X27, p. 6c.

20 [Eastern Jin] Zhu Daosheng, Fahua jing shu, X27, p. 4a.

Daosheng calls the Three Vehicles "concealing traces," while the *Lotus* is "true correctness," the "real one": "In the past, concealing traces in the Three Vehicles, the multitude thought this was it. Now wishing to reveal the real one, showing them the true correctness."[21] From the perspective of expedient means and reality, Daosheng strongly disparages the "expedient," even calling it directly "false," while only the real is "beautiful": "What is called breaking the falseness of the three to accomplish the beauty of the one is called true reality."

2. The relationship between expedient means and reality is closely connected to his concept of "principle" (*li* 理). Daosheng's concept of "principle" is his core concept, a distinctive development of his thought system. Daosheng's "principle" has three characteristics: permanent truth, omnipresence, and uniqueness. Although these three characteristics may not entirely accord with Buddhist scriptures, they align very well with Chinese cultural tradition and the thinking patterns of his contemporaries.[22]

Daosheng similarly interprets the *Lotus* through the concept of principle: "Having said the Three Vehicles are expedient means, now clarify that it is one. The Buddha is the ultimate one; 'one' expresses emergence. If principle could have three, the sage could also emerge as three. But there is no three in principle—only the wonderful one."[23] The One Buddha Vehicle corresponds to principle; both represent uniqueness and the absence of opposition.[24]

3. The Great Vehicle is the Buddha Vehicle. Although Daosheng emphasizes that the Three Vehicles are expedient and false while the One Vehicle is real and beautiful, he does not consider the One Buddha Vehicle and the Great Vehicle among the Three Vehicles to be two different things. Based on the scriptural text, Daosheng clearly states that the Three Vehicles are Śrāvaka, Pratyekabuddha, and Bodhisattva.[25] [26] The so-called Great Vehicle or Bodhisattva Vehicle is "the practice toward buddhahood." The Great Vehicle's theory is broad and deep, capable of liberating from the suffering of saṃsāra; only bodhisattvas can study it.[27]

Why then distinguish between the Bodhisattva Vehicle and the Buddha Vehicle? This comes from the different degrees of realization of "principle" by bodhisattvas and buddhas: "bodhisattvas have not exhausted principle" while "buddhas have completely exhausted principle." Therefore, the unification of the Three into One actually means abandoning the Small Vehicles to enter the Great Vehicle.[28] [29] In sum, Daosheng belongs to the "Three-Vehicle School" rather than the "Four-Vehicle School"—there is no independent Buddha Vehicle apart from the Great Vehicle.

21 [Eastern Jin] Zhu Daosheng, Fahua jing shu, X27, p. 1b.

22 Fu Xinyi, 'Zhu Daosheng chanti chengfo shuo xinlun,' Zhexue yanjiu 2014, no. 6, p. 111.

23 [Eastern Jin] Zhu Daosheng, Fahua jing shu, X27, pp. 4c-5a.

24 [Eastern Jin] Zhu Daosheng, Fahua jing shu, X27, p. 1c. Sometimes Daosheng expresses negation of uniqueness, such as 'Having neither two nor three, the one also departs' (X27, p. 5a). But this expresses the non-oppositional nature of the One Vehicle principle from the negative side.

25 [Eastern Jin] Zhu Daosheng, Fahua jing shu, X27, p. 5b.

26 [Eastern Jin] Zhu Daosheng, Fahua jing shu, X27, p. 3c.

27 [Eastern Jin] Zhu Daosheng, Fahua jing shu, X27, p. 2c.

28 [Eastern Jin] Zhu Daosheng, Fahua jing shu, X27, p. 5b; p. 4b.

29 [Eastern Jin] Zhu Daosheng, Fahua jing shu, X27, p. 1c.

(4) Liu Qiu

Liu Qiu of the Southern Qi, together with more than ten renowned monks, combined the strengths of eight schools to compile the *Annotated Lotus Sūtra*. This book was lost long ago; now we can only see fragmentary passages in Jizang's citations.[30] [31]

Liu Qiu also particularly elaborates on the One Vehicle from the perspective of non-opposition. The One Vehicle is "the ultimate one"; the true Wonderful Dharma does not fall into either extreme: "emptiness and existence are equally exhausted," "emptiness and existence are equally obscured," "neither three nor one."[32] Liu Qiu's approach is similar to Sengrui and Huiguan in interpreting the *Lotus* through emptiness: "nameless and formless—this is the sūtra's ultimate meaning." This has considerable similarity with Jizang.[33]

Overall, these *Lotus* commentators who had close interactions with Kumārajīva, except for Sengrui whose views are relatively ambiguous, can all be classified as the Three-Vehicle School, according relatively well with the original meaning of the scriptural text. "Under the outstanding translator Kumārajīva, most could probably correctly grasp the original meaning of the Sanskrit."[34]

The Early Four-Vehicle School

The emergence of the Four-Vehicle School represents a major turning point in the interpretation of the Parable of the Three Carts. Within the Four-Vehicle interpretation, the distinctively Chinese system of the Perfect Teaching gradually developed. Early proponents of the Four-Vehicle interpretation include Guangzhai Fayun and Nanyue Huisi.

(1) Guangzhai Fayun

Fayun, together with Zhizang (458-522) and Sengmin (467-527), was listed among the Three Great Dharma Masters of the Liang period. Jizang called him "unsurpassed in his time regarding the *Lotus*," "the doctrine of the *Lotus*." He can be regarded as the pinnacle of *Lotus* scholarship during the Northern and Southern Dynasties.[35] [36]

The importance of Fayun lies in his many pioneering views; later commentators, whether agreeing or disagreeing with him, could not bypass him. Both Zhiyi and Jizang critically inherited from Fayun. Moreover, the *Lotus Sūtra Record of Meaning* was written before the translation of the *Lotus Sūtra Treatise*, whereas later commentators were all influenced to varying degrees by that treatise.[37]

30 [Sui] Jizang, Fahua xuanlun, T34, p. 363c.

31 [Sui] Jizang, Fahua xuanlun, T34, pp. 397a-b.

32 [Sui] Jizang, Fahua xuanlun, T34, p. 411b; Fahua youyi, T34, p. 638b.

33 [Sui] Jizang, Fahua xuanlun, T34, p. 381a.

34 [Japan] Hirakawa Akira et al., Hokke shisō (Beijing: Dongfang Chubanshe, 2019), p. 195.

35 [Sui] Jizang, Fahua xuanlun, T34, pp. 363c, 377b.

36 [Sui] Jizang, Fahua xuanlun: 'Guangzhai received the sūtra from Dharma Master Yin of Zhongxing Monastery. Yin was originally from Shouchun, surnamed Zhu. In his youth, he traveled to Pengcheng and received treatise studies from Tandu. Next, he received Lotus studies from Huilong of Kuangshan.' T34, p. 363c.

37 Although Vasubandhu's Fahua lun speaks of the dharma-body, it is closer to the originally pure nirvāṇa

First, the Three Vehicles are "expedient means," the One Vehicle is "true reality," and they correspond to the Tathāgata's two kinds of wisdom. "Having taught today's One Vehicle principle of true reality, this then reveals by contrast that the former Three Vehicles were expedient means."[38] Fayun tends to oppose "expedient means" and "true reality," with a gap between expedient and real concerning whether or not they are ultimate.[39 40]

As Fazang stated: "Dharma Master Yun of Guangzhai Monastery in the Liang period established the teaching of Four Vehicles, as in the *Lotus*—the three carts at the gate are the Three Vehicles; the great white ox-cart bestowed on the crossroads is the Fourth Vehicle. This is because the ox-cart at the gate, like the goat and deer carts, is also not obtained..."[41]

Second, he uses cause and effect to explain the *Lotus* doctrine. The character "wonderful" in the title means the One Vehicle is "wonderful cause and wonderful fruit," while the Three Vehicles represent "former coarse cause and coarse fruit."[42 43] In this way, the One Vehicle has specific content different from the Three Vehicles and a preliminary independent status.

Third, the One Buddha Vehicle stands independently outside the Three Vehicles. This makes Fayun the first "Four-Vehicle" proponent in extant literature and indeed in history.[44 45 46 47]

However, Fayun's Four-Vehicle position is still in an immature form, or can be said to be in a transitional stage between Three Vehicles and Four Vehicles. In sum, under Fayun's conception of expedient means and reality—the Three Vehicles are expedient while the One Vehicle is real—the real One Vehicle cannot be equated with the expedient Great Vehicle among the Three Vehicles. It has independent status and concrete content.[48]

(2) Nanyue Huisi

Nanyue Huisi had deep faith in the *Lotus Sūtra*, as seen in his "Text of Establishing Vows." His specialized work on the *Lotus* is the *Meaning of the Practice of Ease and Joy in the Lotus Sūtra*, developing his understanding of the "Chapter on the Practice of Ease and Joy."

First, he explicitly interprets the One Vehicle through *tathāgatagarbha*. Huisi says: "What

of self-nature in Yogācāra thought, not tathāgatagarbha thought. However, in Northern teachers' interpretations, it was often explained through tathāgatagarbha theory. See Hu Yao, 'Guangzhai Fayun yanjiu,' doctoral dissertation, Sichuan University, 2010; Feng Zixiang, 'Dunhuang yijuan Nanbeichao Lidu fashi Fahua yiji jiejing sixiang yanjiu,' Fayin 2022, no. 2.

38 [Liang] Fayun, Fahua jing yiji 法華經義記 , T33, p. 592a.

39 [Liang] Fayun, Fahua jing yiji: 'The former explains expedient wisdom, which is opening the gate of expedient means; the latter explains real wisdom, which is revealing the characteristic of true reality.' T33, p. 593b.

40 [Liang] Fayun, Fahua jing yiji, T33, p. 593a.

41 [Tang] Fazang, Huayan jing tanxuan ji 華嚴經探玄記 , T35, p. 111b.

42 [Liang] Fayun, Fahua jing yiji, T33, pp. 573b-574a.

43 [Liang] Fayun, Fahua jing yiji, T33, pp. 572c-573a.

44 [Liang] Fayun, Fahua jing yiji, T33, p. 604b.

45 [Liang] Fayun, Fahua jing yiji, T33, p. 619a.

46 The scriptures have two theories of seventh or eighth stage; see Mahāprajñāpāramitā-śāstra, T25, p. 571a; Laṅkāvatāra Sūtra, T16, pp. 489c-490a; Śrīmālā Sūtra, T12, p. 221c. Fayun takes the eighth-stage theory; see Fahua jing yiji, T33, p. 580b.

47 [Liang] Fayun, Fahua jing yiji, T33, p. 621c.

48 [Liang] Fayun, Fahua jing yiji, T33, p. 572c.

is called One Vehicle? All sentient beings, through the *tathāgatagarbha*, are ultimately eternally at ease and joyful." "All sentient beings are fully endowed with the dharma-body treasury, identical with Buddha, without difference."[49 50]

The interpretation of the *Lotus* through *tathāgatagarbha* thought arose early in China but was not accepted by early commentators. Huirui once mentioned a view that treats the *Lotus* as scriptural evidence that all sentient beings can become Buddha, against which Huirui gently argued.[51] None of Kumārajīva's disciples gave a clearly *tathāgatagarbha* interpretation of the *Lotus*; they were closer to the Emptiness School.[52]

But the most important reason for Huisi's interpretation is his own contemplative experience. When Zhiyi had doubts about the *Sutra of Golden Light*'s teaching that "one mind contains ten thousand practices," Huisi said: "Your earlier doubt—this is merely the meaning of the *Prajñāpāramitā*'s sequential approach, not yet the *Lotus*'s perfect and sudden purport. In a past summer, through bitter discipline contemplating this, in a single thought in the later night suddenly all dharmas arose. I have personally realized this; do not trouble yourself with doubt."[53]

Second, based on this foundation, Huisi distinguishes between dull-faculty bodhisattvas and sharp-faculty bodhisattvas. Sharp-faculty bodhisattvas are "not practicing sequentially, nor eliminating afflictions"; "one mind, one study, all fruits are complete together—accomplished simultaneously, not entering sequentially." From the perspective of not relying on sequence, this is called "One Vehicle."[54 55]

Huisi should be considered the true founder of the Four-Vehicle School. Combining his own contemplative experience, he further polarized the relationship between expedient means and reality, distinguishing sharp-faculty bodhisattvas from dull-faculty bodhisattvas. Through Zhiyi's propagation, later Four-Vehicle proponents abandoned Fayun's path and continued forward on the path Huisi had opened.

Concluding Remarks

In summary, the earliest Lotus commentators were all Three-Vehicle proponents, very likely because they mostly studied under Kumārajīva and had a relatively accurate understanding of the Lotus's original meaning. The Four-Vehicle School emerged because of temporal distance; through discovering the interpretive space in the Chinese text, they elaborated their own Buddhist thought. The interpretive modes of the Three-Vehicle and Four-Vehicle Schools might be compared to "I annotate the classics" versus "the classics annotate me."

Fayun represents the crucial turning point from Three Vehicles to Four Vehicles. He strengthened the opposition between expedient means and reality and endowed them with concrete and detailed content, thereby preliminarily establishing the One Vehicle's

49 [Chen] Huisi, Fahua jing anlexing yi 法華經安樂行義 , T46, pp. 698a-b.

50 [Chen] Huisi, Zhufa wuzheng sanmei famen 諸法無諍三昧法門 , T46, p. 632b.

51 [Liang] Sengyou, ed., Chu sanzang ji ji, T55, pp. 42a-b.

52 [Liang] Fayun, Fahua jing yiji, T33, pp. 603b-c, 621b-c.

53 [Tang] Daoxuan, Xu gaoseng zhuan 續高僧傳 , T50, pp. 563a-b.

54 [Chen] Huisi, Fahua jing anlexing yi, T46, p. 698a.

55 [Chen] Huisi, Fahua jing anlexing yi, T46, p. 698c.

independent status. Building on this foundation, Huisi used his own contemplative experience and his clearly articulated tathāgatagarbha theory to open the second path of the Four-Vehicle School.

Why did the transformation from Three Vehicles to Four Vehicles occur? First, parables themselves have multiple interpretive possibilities. Second, endowing the One Vehicle with independent status and concrete content different from the Three Vehicles. If we trace back to Indian Mahāyāna thought, the "greatness" of Mahāyāna has two interpretive directions: inclusive greatness and excellence greatness. As the *Mahāprajñāpāramitā-śāstra* says: "Mahāyāna is vast; all vehicles and paths enter Mahāyāna... The great ocean can receive all streams because of its vastness."[56]

Third, the continual polarization of the relationship between expedient means and reality. Fayun's interpretive consciousness was deeply influenced by the phrase "opening the gate of expedient means to reveal the characteristic of true reality," and "clearly, consciously or unconsciously, strengthened the opposition between 'opening the gate of expedient means' and 'revealing the characteristic of true reality.'"[57] [58]

From a broader perspective, the *Lotus Sūtra* phrases "opening the gate of expedient means to reveal the characteristic of true reality" and "straightforwardly abandoning expedient means" have not only Buddhist doctrinal significance but also hermeneutical significance. "In the era of Wei-Jin and Northern-Southern Dynasties Chinese Buddhist scriptural hermeneutics, especially in *Lotus Sūtra* hermeneutics, this translated phrase exerted unimaginably important influential effects."[59]

The continual polarization of the relationship between expedient means and reality influenced religious practice: naturally, one would choose the good and beautiful real rather than the artificial and false expedient, choose quick attainment of buddhahood rather than kalpas of cultivation, choose the perfect and sudden rather than the separated and sequential. Thus, in the dogmatized sectarian Buddhism that followed, there gradually formed the tendency to disdain the Three Vehicles and exclusively venerate the One Vehicle that is "purely perfect and uniquely wonderful"—leading Chinese Buddhism to increasingly lose the six pāramitās and ten thousand practices of the bodhisattva path along with the stages of cultivation.

But reality and expedient means were not originally opposed. In Kumārajīva's translation, "reality" refers to "the bodhi realized by the Buddha," belonging to realized dharma, while skillful means "is precisely what leads to 'reality,'" belonging to taught dharma—no opposition exists between them.

Perhaps we should return to the reminder of the *Lotus* translator Kumārajīva: "The *Lotus Sūtra* is the secret treasury of all Buddhas; one cannot use this meaning to challenge other sūtras. If one fixedly adheres to the *Lotus* as definitive, the Śrāvaka Tripiṭaka and other Mahāyāna sūtras would be set aside unused... If so, only the *Lotus* alone would be trustworthy; all other sūtras would be false. Therefore, one should not be attached to one sūtra and

56 [Later Qin] Kumārajīva, trans., Mahāprajñāpāramitā-śāstra 大智度論 , T25, p. 86a.

57 Cheng Gongrang, Fodian Hanyi, lijie yu quanshi yanjiu 佛典漢譯、理解與詮釋研究 (Beijing: Zhongguo Shehui Kexue Chubanshe, 2017), p. 347.

58 Cheng Gongrang, Fodian Hanyi, lijie yu quanshi yanjiu, pp. 346-347.

59 Cheng Gongrang, Fodian Hanyi, lijie yu quanshi yanjiu, p. 331.

disbelieve all other sūtras and dharmas."[60]

If one is attached to one scripture as supreme, considering the teachings of other scriptures non-ultimate, even setting aside other scriptures, this ultimately leads to the predicament of self-enclosure and self-exaltation without self-awareness. Nearly 1,600 years later, Kumārajīva's voice remains powerfully resonant. This is also the insight that the evolution of the Parable of the Three Carts interpretations brings us.

Yan Jun is an Assistant Research Fellow at the Center for Judaic and Interreligious Studies and the Department of Religious Studies, School of Philosophy and Social Development, Shandong University. His research interests include Yogācāra studies, Buddhist logic (Hetuvidyā), Madhyamaka philosophy of the middle period, Buddhist hermeneutics, and the relationships among Confucianism, Buddhism, and Daoism.

60 [Later Qin] Kumārajīva, Jiumoluoshi fashi dayi 鳩摩羅什法師大義 , T45, p. 133b.

Journal of Language, Media and Society
Vol. 1, No. 1 https://doi.org/10.66056/jlms.2026.v1i1.005

Carter & Co. PUBLISHING

FRACTURED SILENCE: PROPERTY ANXIETY, INTERNAL DIVISION, AND THE SELF-DISCIPLINING MIDDLE CLASS IN POST-LOCKDOWN SHANGHAI

JINPU WANG

This paper examines the political behavior of China's urban middle class during and after the 2022 Shanghai COVID-19 lockdown through a longitudinal digital ethnography of WeChat groups in a residential compound from 2022 to 2025. Extending an earlier study of "bounded resistance" during the lockdown crisis, this research tracks the evolution of community dynamics through the subsequent economic downturn. Contrary to theoretical expectations that middle-class expansion leads to political liberalization or that economic grievances catalyze collective resistance, the study finds that Shanghai's middle class neither pushed for political change during prosperity nor mobilized against the state during decline. Instead, residents actively participated in maintaining social order through three mechanisms, namely surveillance-induced self-censorship, peer discipline against "defectors" who emigrated or sold property below market price, and displacement of economic anxieties onto inter-stratum conflicts between homeowners and renters, locals and migrants. I argue that China's "middle class" functions in this context not as a unified political actor but as a fragmented aggregate defined by property ownership, without the shared interests or collective capacity that class-based theories assume. The concept of "cynical entrapment," a condition where both exit and voice are systematically blocked, leaving neither genuine loyalty nor viable alternatives, helps explain the paradox of widespread grievance without mobilization. These findings contribute to debates on authoritarian resilience by demonstrating how the Chinese state achieves stability through minimal intervention, leveraging social divisions and property-based identities to generate self-policing communities.

KEYWORDS: middle class, China, state-society relations, authoritarian resilience, self-discipline, property anxiety, digital ethnography, Shanghai

Introduction

In the autumn of 2024, a former resident of my Shanghai compound posted a sardonic comment in our community WeChat group. She was a software engineer who had emigrated to Canada two years earlier. The discussion had turned to declining property values in the neighborhood, a topic that had become increasingly fraught as housing prices continued their downward slide. "Glad I got out when I did," she wrote. "The whole market is heading south." Within minutes, she was besieged by hostile responses from current residents.

"If you've left, why are you still here stirring up trouble?"

© 2026 Carter & Co.

"Don't forget who made your wealth possible in the first place. Show some gratitude instead of gloating."

"You're exactly the kind of person dragging down our property values. Selling cheap and running away. Traitors have no right to comment on our affairs."

The barrage continued for nearly an hour, with dozens of residents piling on. The emigrant attempted a few defensive replies. She had sold at market price; she still cared about her former community. But she was outnumbered and outgunned. By the next day, her messages had stopped. Her account had been removed from the group, whether voluntarily or by the administrator. She never posted again in other group chats either.

This incident crystallized a puzzle I had been observing for three years. Why does China's urban middle class not only refrain from challenging state authority but also actively police itself against any discourse that might destabilize the existing order? The emigrant had not criticized the government. She had merely noted an objective fact about property prices. Her crime was making visible the possibility of exit, threatening the collective fiction that staying put was the only rational choice. Her silencing was not ordered by the state. Rather, it was executed by her former neighbors.

This paper asks why China's expanding middle class often fails, and likely continues to fail to become a force for sociopolitical change during both economic prosperity and economic decline. The question confounds two major theoretical traditions. Modernization theory, from Lipset (1959) through Inglehart and Welzel (2005), predicts that economic development and middle-class expansion will generate demands for political liberalization. Relative deprivation theory (Gurr 1970; Walker and Smith 2002) predicts that unmet expectations breed grievances capable of motivating collective action. The Chinese case defies both. During the decades of rapid growth, the middle class remained politically quiescent, accepting an implicit exchange of political freedom for stability and prosperity (Perry 2008). When the 2022 lockdown and subsequent economic downturn visibly broke that contract, the predicted mobilization still failed to materialize.

I argue that the Chinese middle class is not merely a passive beneficiary of authoritarian stability but an active participant in maintaining it. Drawing on a longitudinal digital ethnography of WeChat groups in a middle-class residential compound in Shanghai from 2022 to 2025, I document three interconnected mechanisms through which middle-class communities constitute themselves as self-disciplining units. First, the ambient presence of state actors in community spaces creates a panopticon effect that induces generalized self-censorship. Second, community members actively punish those who emigrate, sell property below market price, or express exit-oriented discourse, blocking the informational function of exit and maintaining the fiction that alternatives do not exist. Third, internal divisions between homeowners and renters, Shanghai natives and migrants, and those who acquired property through different channels generate horizontal conflicts that absorb the energy that might otherwise fuel collective grievance against the state.

The result is what I term "fractured silence." Widespread grievances fail to translate into collective voice while exit options are discursively suppressed even as they are privately pursued. This is not the loyal silence of citizens who trust their government, nor the fearful silence of those cowed by repression. It is a cynical silence maintained through mutual surveillance and horizontal hostility. I conceptualize this condition as "cynical entrapment," a modification of Hirschman's (1970) Exit-Voice-Loyalty framework in which exit is constrained by property anchors and capital controls, voice is suppressed by surveillance and peer discipline, and what remains is neither loyalty nor any viable alternative.

This study extends my earlier research on the same compound during the 2022 Shanghai lockdown (Wang and Xiang 2025), which documented what I termed "bounded resistance," fierce criticism of local officials combined with restraint in challenging the central state's zero-COVID policy. Where that study asked why resistance remained bounded during crisis, the present study asks what happens when the crisis passes but grievances persist. The answer is not the emergence of new forms of resistance but the deepening of self-discipline, a shift from bounded resistance to fractured silence.

The paper makes three theoretical contributions. First, it extends the critique of modernization theory by showing that the Chinese middle class actively undermines the conditions for its own political agency through internal fragmentation and property-based identity structures. Second, it complicates relative deprivation theory by demonstrating how economic grievances can be absorbed by horizontal conflicts between homeowners and renters, locals and migrants, stayers and leavers, rather than channeled into collective challenge against the state. Third, it offers a revised Exit-Voice-Loyalty framework for authoritarian contexts in which neither exit, voice, nor loyalty adequately describes the political condition of subjects who remain in a system they no longer believe in, without alternatives they dare pursue.

The empirical setting is post-lockdown urban China, a period economists have termed "long COVID" for its persistent economic malaise. In particular, the property market entered a prolonged crisis. Prices in major cities fell 20 to 50 percent from their peaks; mass-scale developers like Evergrande and Country Garden defaulted on hundreds of billions in debt; consumer confidence collapsed to historic lows. For Shanghai's middle class, roughly 70 percent of whose household wealth is concentrated in real estate, this represented an existential threat to their achieved status. Yet the predicted political consequences did not materialize. The "white paper protests" of November 2022, while historically "significant" as some China observers praised, were very limited in scale and quickly dissolved on their own, without even much state suppression. What followed was not sustained mobilization but what Ong (2023) calls an "epidemic of mistrust," meaning pervasive disillusionment that found expression in emigration, withdrawal, and cynicism rather than collective action.

Methodologically, this study demonstrates the value of longitudinal digital ethnography for studying political behavior under authoritarianism. Following Hine's (2015) "embedded, embodied, everyday" framework, I treat WeChat groups as extensions of physical community life rather than separate virtual spaces. The three-year design reveals patterns invisible in cross-sectional analysis, including the gradual deepening of silence and the shifting targets of frustration. My position as a non-local renter, the lowest status in the community hierarchy, gave me firsthand experience of the tensions I analyze.

A note on terminology. I use "middle class" to refer to urban residents with middle-range incomes, property ownership or substantial rental expenditure, and white-collar occupations, the group Chinese discourse terms "*中产阶级*." However, a central argument of this paper is that this group is better understood as a property-defined stratum than a class in the Weberian or Marxian sense. Its members share a similar economic position but lack shared interests, common identity, or demonstrated capacity for collective action. Under the conditions documented here, this fragmentation is constitutive of their political quiescence. The term is therefore used descriptively rather than analytically.

Literature Review

This study engages three interconnected bodies of scholarship. The first concerns the middle class, authoritarian resilience, and community governance. The second addresses digital surveillance and the mechanisms of self-censorship. The third examines the relationship between property, class identity, and the exit-voice-loyalty framework. Each offers partial insights into middle-class political attitude and behavior. Together, they help frame the empirical puzzle of why China's urban middle class neither demands liberalization during prosperity nor mobilizes during decline.

Middle Class, Authoritarian Resilience, and Community Governance

Scholars have long debated whether the expansion of the middle class promotes democratization. Lipset's (1959) foundational hypothesis linked economic development to political liberalization, and Moore's (1966) complementary thesis identified class coalitions as determinants of regime type. Subsequent work further complicated this picture. Przeworski and colleagues (2000) showed that economic development helps democracies survive but does not cause them to emerge. Rueschemeyer, Stephens, and Stephens (1992) argued that capitalism is associated with democracy not because of the bourgeoisie attributes but because it strengthens subordinate groups. The bourgeoisie, they found, has often opposed democracy out of fear of redistribution.

China-focused research has documented the failure of these expectations with evidence from various settings in the country. Chen's (2013) survey-based study found that China's middle class, especially state-sector employees, is more supportive of the party-state but less supportive of democratic values than lower classes. Only 11 percent of state-sector middle class showed high democratic support, compared with 49 percent in the private sector. Chen introduced the concept of "contingent democratic supporters" to capture how middle-class political orientation depends on proximity to the state and perceived economic wellbeing rather than principled commitment. Dickson's (2016) work documented surprisingly high levels of popular support for the CCP, finding that citizens prefer incremental change within the existing framework over systemic transformation. Wright's (2010) concept of "structural acceptance" showed that citizens tolerate authoritarianism not from coercion but from calculated self-interest shaped by state-led development, late industrialization, and socialist legacies.

Nathan's (2003) influential concept of "authoritarian resilience" identified institutional adaptations that distinguish China from personalist dictatorships prone to collapse. O'Brien and Li's (2006) work on "rightful resistance" showed how citizens frame grievances within system-affirming boundaries, exploiting central-local divisions while accepting the regime's legitimacy. This helps explain the pattern observed during the Shanghai lockdown, where residents directed criticism at local officials while affirming the central state's good intentions. Yet rightful resistance assumes available discursive space for complaint. What happens when that space contracts?

Tomba's (2014) analysis brought the question to the neighborhood level most relevant to this study. Based on fieldwork in multiple Chinese cities, Tomba argued that residential neighborhoods are sites of "intense governing" where the state maintains legitimacy through differentiated strategies. Middle-class neighborhoods receive greater autonomy in exchange

for social order maintenance, producing "contained contention" where conflicts remain within gated structures. His concept of "social clustering" captures how physical and social segregation keeps conflicts localized. Meanwhile, the grid management system (*网格化管理*), expanded dramatically under Xi Jinping, deploys millions of grid workers who simultaneously serve as eyes of the state and providers of community services, making resistance costly while creating dependence (Mittelstaedt 2022; Wang and Xiang 2025).

This thread of literature establishes that the Chinese middle class does not conform to modernization theory's expectations and that the state has developed sophisticated mechanisms of community governance. What it less adequately explains is the mechanisms behind why the middle class might actively participate in maintaining authoritarian stability, not merely failing to demand change but working to suppress those who do. The present study aims to address this gap.

Digital Surveillance and Self-Censorship

Research on Chinese digital governance has transformed the understanding of how authoritarian regimes leverage technology for social control. A series of work by King, Pan, and Roberts established that the government does not primarily censor criticism of leaders or policies. Instead, censorship targets posts that could catalyze collective action regardless of their political valence (King, Pan, and Roberts 2013). Another work revealed that the "50-cent party" floods social media with cheerful, distracting content rather than arguing with critics, producing an estimated 448 million fabricated posts annually (King, Pan, and Roberts 2017). The strategy is distraction, not persuasion. Roberts' (2018) synthesis identified three mechanisms of censorship, namely fear, friction, and flooding, and documented how the Great Firewall creates "social segmentation" between a skeptical, tech-savvy class and the broader public, preventing coordination between elite opinion-leaders and mass audiences.

Huang's (2015) research on propaganda as signaling added a crucial dimension. Survey data showed that those exposed to more ideological education did not hold more positive views of the government but were more likely to believe the regime is strong and less willing to participate in dissent. Propaganda works not through persuasion but by demonstrating organizational capacity and reach. This signaling function helps explain how sparse enforcement can produce widespread compliance. The occasional reminder of state presence activates self-censorship far beyond what direct monitoring would require.

Research on self-censorship documents the mechanism most central to this study. Robinson and Tannenberg's (2019) list experiments found that 24.5 to 26.5 percentage points more individuals express regime support through direct questioning than through indirect methods, indicating massive preference falsification. Critically, middle-class urban residents self-censored more than rural residents, suggesting that higher stakes and greater awareness of surveillance produce more cautious speech. Yang's (2025) research on the "normalization of censorship" showed that when censorship expands to include non-political content, citizens become desensitized and political censorship provokes less backlash.

Regarding sampling source, most existing studies focus on public platforms such as Weibo and WeChat Moments rather than semi-private spaces like community WeChat groups. The present study extends analysis to these spaces and finds that state censorship operates as a background condition enabling a more pervasive system of social self-censorship. The key mechanism is not fear of state punishment but anticipatory conformity driven by peer pressure.

Residents self-censor because they fear social marginalization, not necessarily immediate sanction from the state.

Property, Exit, and Class Identity

The centrality of property to Chinese middle-class identity requires engagement with scholarship on housing, stratification order, and political behavior. The 1998 housing reform transformed housing from a socialist welfare benefit into a capitalized private asset, with urban homeownership rising from roughly 35 percent in 1995 to over 85 percent by 2010 (Davis 2003). Xie and Jin (2015) found that housing assets account for over 70 percent of total household wealth in China and served as the main driver of drastically rising wealth inequalities in the recent two decades. This concentration creates what I term a "property anchor," financial exposure so substantial that it binds the new urban middle-class to the existing system regardless of their political preferences. Cai's (2005) research found that homeowners are "moderate" in their activism, pursuing narrow economic interests such as disputing management fees or protesting construction quality rather than systemic political change.

Hirschman's (1970) Exit-Voice-Loyalty framework provides essential conceptual tools for understanding responses to economic decline. Members of declining organizations choose between exit and voice, with loyalty moderating this choice. Clark, Golder, and Golder (2017) reformulated this for authoritarian contexts, arguing that citizens choose loyalty not from special attachment but because they are powerless to do otherwise when lacking credible exit threats. In China, the middle class lacks credible exit options (legal emigration is difficult and costly), and the government does not depend on this class specifically, unlike capital-owning elites. Recent scholarship on Chinese emigration confirms these constraints. Chau and Gherghina (2024) found that wealthy Chinese emigrants' loyalty was "economically conditional," ending when Xi Jinping's policies threatened their economic interests and personal safety. The post-2022 "润学" (runology) phenomenon and the viral discussion of emigration strategies represent unprecedented public engagement with exit as a political response, though actual emigration remains limited to those with the resources to execute it.

This study proposes a modification to Hirschman for the Chinese middle-class context. I find that exit behavior is more prevalent than popularly perceived but discursively suppressed. People do emigrate, but acknowledging emigration as a legitimate response to the political-economic reality often triggers community sanction. Voice is constrained by surveillance and peer discipline. What remains is not loyalty in Hirschman's sense but what I term "cynical entrapment," a condition of continued participation despite evident disillusionment, a shared understanding that alternatives are blocked, and active suppression of discourse about those alternatives. The community attacks recent emigrants not because emigration threatens the state but because it threatens the shared understanding that staying is the only reasonable course of action.

Together, these three literatures frame the present study's contribution. Modernization theory and its critics explain why the middle class fails to demand liberalization during prosperity. State-society and digital governance scholarship explains how the state penetrates and monitors communities. Exit-Voice-Loyalty theory provides a framework for understanding responses to decline. This study brings these perspectives together by documenting how middle-class communities internalize and reproduce control functions, generating stability

from below through peer surveillance, social sanction, and the displacement of grievances onto horizontal conflicts. It differs from the COVID lockdown resistance literature by examining not protest but the sedimentation of silence after protest subsides. It also differs from censorship scholarship by analyzing social enforcement in semi-private community groups rather than platform-level content moderation. Furthermore, it is parallel to the longstanding homeowner activism studies on "not-in-my-backyard" type of collective resistance by showing property not only as a basis for mobilization but as an anchor that fragments collective capacity. And my findings contrast existing applications of the Exit-Voice-Loyalty framework by documenting how community members themselves delegitimize exit, blocking its informational function from below.

Research Context and Methods

This study employs longitudinal digital ethnography to examine middle-class political behavior in a Shanghai residential compound from March 2022 to December 2025. Following Hine's (2015) framework of "embedded, embodied, everyday" internet research, I treat WeChat groups as extensions of physical community life rather than separate virtual spaces. The longitudinal design enables analysis of how community dynamics evolved from the acute crisis of the 2022 lockdown through the prolonged economic downturn that followed.

The Field Site

The research site is a middle-class residential compound (小区) in Shanghai's Baoshan District, comprising approximately 2,000 households across multiple high-rise buildings. Developed in the early 2010s as part of Shanghai's urban expansion, the compound replaced semi-rural land on the city's northern periphery. Its location, accessible to the city center by metro but distinctly peripheral, shapes its demographic composition and social dynamics.

The resident population falls into three groups with divergent interests and identities. The first, roughly 20 percent of households, consists of Shanghai natives who acquired apartments through demolition compensation (拆迁户). These families became property owners without purchasing property, also middle-class in housing wealth without the educational or occupational credentials typically associated with that status. Many work in service or manual occupations. The second group, comprising the majority of owner-occupants, consists of migrants from other provinces who purchased apartments at market price. Many hold university degrees and work in professional positions. Most carry substantial mortgage debt on apartments now worth 20 to 30 percent less than what they paid. The third group consists of renters, also predominantly migrants with professional profiles, who participate in community WeChat groups but occupy an ambiguous position, physically present yet lacking the property stake that defines community membership in the eyes of many homeowners, especially the Shanghai natives.

This tripartite structure generates persistent tensions. Shanghai natives who acquired property through urban sprawl claim superior status by virtue of local identity (本地人), sometimes referring to highly educated migrants as "乡下人" (country people, implying "peasant" background) and inverting conventional status hierarchies. Migrants who purchased at market price resent both the "unearned" property of demolition recipients and the presence of renters who "have no stake" in property values. Renters experience exclusion from full

community membership while sometimes privately benefiting from the property decline that devastates owners. These divisions, as I will show, absorb much of the energy that might otherwise fuel collective grievance against the state.

The compound experienced the 2022 Shanghai lockdown with particular intensity, entering "silent management" (*静态管理*) in late March and remaining under various restrictions until early June. During approximately eight weeks of confinement, WeChat groups became essential infrastructure for survival and the primary venue for expressing frustration. My earlier study (Wang and Xiang 2025) documented the "bounded resistance" that emerged during this period. The present study extends observation through December 2025, encompassing the lockdown's aftermath and the prolonged economic downturn. Property values in the compound fell roughly 25 percent from their 2021 peak. The transformation from the intense collective emotion of the lockdown to the fractured silence of 2024–2025 constitutes this study's central empirical puzzle.

Data Collection

The primary data source is participant observation in eight WeChat groups associated with the compound. Three are official homeowners' association (*业委会*) groups with 200 to 500 members, including property management representatives, neighborhood committee officials (*居委会*), and grid workers (*网格员*). Five are informal networks that emerged during the lockdown for group purchasing (*团 购 群*), ranging from 50 to 300 members with less official presence, though the boundary between official and informal groups is porous.

I joined these groups as a resident, having rented an apartment in the compound between 2021 and 2022. My membership predates any research intention. I became a participant-observer as the lockdown transformed these groups from mundane coordination tools into sites of intense social and political significance. Following Kozinets' (2019) netnographic framework, the analysis draws primarily on archival data, the flow of messages, discussions, and conflicts constituting everyday group life, supplemented by fieldnotes recording my observations and interpretations. Data collection involved systematic archiving through screenshots and exports organized chronologically, with complete threads captured for significant incidents. The resulting archive comprises approximately 3,000 screenshots and text exports. Supplementary data includes informal conversations with neighbors in elevators, courtyards, and the community's small commercial area, as well as monitoring of broader social media discussion on Weibo and Xiaohongshu.

Analytical Approach

Analysis followed Braun and Clarke's (2006, 2022) reflexive thematic analysis, an approach that treats themes not as entities "emerging" from data but as patterns actively constructed through interpretive engagement. The process involved familiarization through repeated reading, initial coding, theme searching and reviewing, and iterative refinement as interpretation developed. Initial coding identified over 200 codes capturing discourse features, interaction patterns, and expressed sentiments, organized using MAXQDA software for its handling of Chinese-language text.

The longitudinal dimension was analytically central. I compared discourse patterns during the lockdown (March–June 2022), the immediate aftermath (late 2022–2023), and the prolonged downturn (2024–2025), in order to identify both continuities and transformations. Three master themes emerged from this process, corresponding to the mechanisms documented in the findings. All primary data is in Mandarin Chinese, and I conducted analysis in the source language to preserve nuance. Quotations were translated with attention to conveying meaning rather than literal equivalence. Certain terms resist translation. " *负能量* " (negative energy) carries connotations of moral failing absent from the English, and " *润* " (run, as in emigrate) puns on a character meaning "moist" or "profitable" in ways that disappear in translation. I have preserved Chinese terms where English equivalents would lose significant meaning. The excerpts presented in this paper were selected as representative instances of patterns that recurred across groups and time periods, not as isolated incidents. I also searched systematically for counterexamples, including moments where dissent was sustained without sanction, where exit was discussed supportively, or where residents formed cross-cleavage alliances around shared grievances. Such moments were rare and typically short-lived, which itself constitutes evidence for the mechanisms I describe.

Researcher Positionality

My position as a non-local renter (*外地租户*) places me at the bottom of the community's informal status hierarchy. I am a migrant, not a Shanghai native. A renter, not an owner. A relative newcomer, not a long-term resident. I experience firsthand the condescension that Shanghai natives sometimes direct at migrants, the suspicion that homeowners harbor toward those without property stakes, and the ambient awareness of surveillance that constrains my own speech. When residents complained about "outsiders who don't understand Shanghai" or suggested that "renters have no right to comment on property issues," I was implicitly included.

This marginal position offers analytical advantages alongside its limitations. I am not invested in defending property values or community reputation, and my outsider status made me less likely to be perceived as a threat. At the same time, my marginality limits access to certain insider knowledge, including homeowners' association meetings and private discussions among long-term residents. Following Dwyer and Buckle's (2009) framework of the "space between" insider and outsider positions, I occupied a liminal status, close enough to understand context and nuance but distant enough to perceive patterns that full insiders might take for granted. I should also acknowledge that I am critical of the Chinese government's pandemic response and its broader authoritarian trajectory. These dispositions inevitably shape my analysis. I have attempted to discipline them through systematic attention to evidence and alternative interpretations, but I do not claim value-free observation.

Ethical Considerations

Research in authoritarian contexts on politically sensitive topics requires ethical protocols that exceed standard requirements. Following the Association of Internet Researchers' guidelines (Franzke et al. 2020), I approached ethics as a continuous process rather than a one-time determination. The question of informed consent in community WeChat groups presents genuine dilemmas. Individual consent from hundreds of group members was neither practical nor, given the sensitive context, desirable. Announcing a research presence might have altered the behavior I sought to observe and, more seriously, might have endangered participants by associating them with research on politically sensitive topics. Following precedent in

WeChat research (Moffa and Di Gregorio 2023; Wang and Xiang 2025) and guidance from my institutional review board, I treated group observation as analogous to observation of semi-public behavior. Specific quotes used in work intended for publication were anonymized and, where feasible, verified with the quoted individuals.

I have removed all identifying information from quoted material, including names, apartment numbers, specific dates, and distinctive phrasing. The compound is identified only by district and general characteristics. Certain non-essential details have been slightly altered to prevent identification while preserving analytical accuracy. I have also taken precautions regarding data security, including encrypted storage and backup outside the Chinese jurisdiction. Finally, I recognize the risk of reproducing orientalist frameworks when writing about Chinese citizens for primarily Western academic audiences. I have attempted to present community members as agents navigating difficult circumstances rather than passive victims of authoritarian control, recognizing their choices as rational responses to structural constraints that would shape behavior in any similarly positioned population.

Findings

The following analysis documents three interconnected mechanisms through which middle-class residents of the Shanghai compound maintained political quiescence during the 2022–2025 observation period. These mechanisms operated in concert to constitute a self-disciplining community in which the state established the conditions for social control but residents performed the daily work of enforcement. I present each with attention to temporal evolution, showing how community dynamics shifted from the acute crisis of 2022 through the prolonged downturn of 2024–2025.

The Production of Silence

The most striking finding of this study concerns the progressive deepening of silence in community discourse over the three-year observation period. This silence was not imposed directly by state censorship. No official ever deleted messages or warned residents about their speech in the groups I observed. Rather, it emerged through the internalization of surveillance logics and, crucially, through peer enforcement of discursive boundaries. The trajectory moved from cautious complaint during the lockdown, through confused withdrawal in the aftermath, to comprehensive self-censorship by 2024–2025. Understanding this trajectory requires attention to both the state's ambient presence and the community's active role in policing itself.

State presence in the WeChat groups was simultaneously known and backgrounded, an open secret that structured interaction without dominating it. In the three homeowners' association groups, representatives from the neighborhood committee (居 委 会) were identified members with their official roles visible in their WeChat profiles. Grid workers (网格员) participated in at least two groups, occasionally posting policy announcements or responding to complaints. During the 2022 lockdown, their presence had been intensive. By 2024 their participation had become sporadic, limited largely to official notices, but their membership remained visible in group member lists.

More significant than these identified officials was the ambient uncertainty about who else might be watching. Residents speculated openly, though carefully, about which neighbors might report conversations to authorities. In October 2023, following a discussion that had drifted toward criticism of economic policy, one resident posted.

"Everyone be careful what you say. There are all kinds of people in this group. Watch out for screenshots and reports."

HOA Group 1, October 2023

The warning invoked a generalized possibility, not specific knowledge of informants. Another resident responded:

"Right, I have a friend who got a visit from the police station just because of a few things she said in a group chat. These days, who knows anyone's real identity?"

Whether this story was true, exaggerated, or apocryphal matters less than its function. It established that surveillance was not merely possible but had consequences, and that anyone might be its agent.

This configuration extends Foucault's (1977) panopticon into digital space. The grid worker's visible presence functioned as a reminder that someone might be watching, producing self-regulation regardless of whether surveillance was active. The result was preemptive self-censorship exceeding what direct monitoring would require. Residents edited themselves not because they knew they were being watched but because they could not know they were not.

This ambient surveillance deepened over time. The longitudinal dimension of this study reveals how silence deepened through three distinct phases.

Phase 1. Bounded Resistance (March–June 2022). During the lockdown the WeChat groups were sites of intense activity, with hundreds of messages daily at peak periods. Residents coordinated group purchases, shared information, and vented frustrations about supply shortages and policy confusion. Criticism was fierce but carefully calibrated, targeting implementation failures while affirming the legitimacy of pandemic control itself. This pattern, which I documented earlier as "bounded resistance" (Wang and Xiang 2025), followed the template of "rightful resistance" (O'Brien and Li 2006). A characteristic exchange from April 2022 illustrates the dynamic. After three days without vegetable delivery, residents erupted in complaint.

"The government promised to guarantee supplies. Where are they? Three days now, what are our elderly and children supposed to eat? What kind of execution is this?"

Group Purchase Group 2, April 12, 2022

The complaint targeted "execution," not policy. When another resident began to question the lockdown's rationale more fundamentally ("Is this really about saving lives or saving face?"), he was quickly corrected.

"Don't go there, it's pointless. The state's intentions are definitely good; the problem is the people implementing the policy below. What we need to solve is the immediate problem, not debate whether policy is right or wrong."

Response in the same thread, April 12, 2022

The framing "国家的出发点肯定是好的" ("the state's intentions are definitely good") operated as a formula, perhaps sincere, perhaps protective camouflage, that contained critique within acceptable bounds. Systemic critique might catalyze collective action; implementation complaints could not.

Phase 2. Confused Withdrawal (Late 2022–2023). The abrupt policy reversal of December 2022, when all COVID restrictions were abandoned virtually overnight, created a period of disoriented silence. The sudden shift contradicted everything residents had been told about

the necessity of zero-COVID. Some expressed bewilderment.

"Just like that, it's over? Then what were the past two months for, locked in our homes? How many businesses destroyed, how many elderly passed without treatment… Now the virus isn't scary anymore? Then what about before?"

HOA Group 2, December 8, 2022

This message received no responses. The silence was eloquent. Engaging with the contradiction was understood as dangerous territory. Over the following months, political discussion largely disappeared from the groups. Activity declined and conversations reverted to pre-lockdown routines, including property management complaints, school enrollment questions, and restaurant recommendations. The lockdown trauma was not processed collectively. It was simply not mentioned. When I asked a neighbor about this silence in early 2023, she replied, "What's the point of talking about that? Can't change anything, just asking for trouble. Look forward." This "looking forward" represented a collective decision to foreclose discussion of the immediate past, enforced not by authorities but by residents themselves.

Phase 3. Cynical Resignation (2024–2025). By 2024 the atmosphere had shifted from confused withdrawal to something I can only describe as cynical resignation. The economic downturn was undeniable. Property values in the compound had fallen approximately 25 percent from their 2021 peak. Several residents had lost jobs or taken salary cuts. Small businesses in the neighborhood had closed. Yet discussion of these conditions remained oblique, carefully depoliticized, and frequently suppressed.

Economic difficulties were acknowledged but framed as natural phenomena rather than policy consequences. In a September 2024 discussion of falling property values, one resident observed.

"This is just the macro environment; the whole world is adjusting. All we can do is hold steady, don't panic-sell (properties). The (real estate) market will come back eventually."

HOA Group 1, September 2024

The framing ("macro environment," "global adjustment," "the market") evacuated agency and causation. The property crisis was presented as a natural disaster to be weathered, not a consequence of policy choices that might be criticized. When another resident tentatively suggested, "Policy keeps changing, who dares buy property?", the response was immediate.

"Talking about policy is pointless; we have no say. Better to think about how to improve our compound's environment, whether we can lower the property management fee. These are things we can actually influence."

Response in the same thread, September 2024

The move was characteristic. From structural cause to manageable symptom. From political critique to practical complaint. Property management, endlessly criticized for poor service and high fees, became the safe target for frustrations that could not be directed at their actual sources. Complaining about the property company was acceptable, even cathartic. It posed no political risk and changed nothing.

Beyond these individual patterns of self-censorship, the most important finding regarding silence was its enforcement not primarily by state actors but by fellow residents. When someone spoke "out of turn," venturing too close to political critique or expressing excessive pessimism, correction came swiftly from peers. This horizontal discipline operated through several mechanisms.

Rapid redirection. When conversations drifted toward sensitive territory, residents intervened to change the subject. In March 2024, a discussion of youth unemployment touched on the broader economic situation.

"It's so hard for young people to find jobs now. My nephew graduated from a 985 (note: "985" is a state-endorsed category of top-tier universities in China) university, sent hundreds of resumes, and hasn't gotten a single interview.

"The whole economy is failing. Used to be that studying hard meant a way out, now education is useless too. This generation really…"

"We've gone off topic. Let's get back to compound matters. What about the parking space problem? Last meeting they said they'd re-draw the lines, any news?"

HOA Group 1, March 2024

The intervention "话题扯远了" ("we've gone off topic") was formally about group relevance but functionally about political risk. The developing critique that "the whole economy is failing" was stopped before it could become explicit. The redirection was itself a message. Everyone understood this was dangerous territory. Return to safe ground.

Silencing through non-response. A subtler mechanism was simply ignoring transgressive posts. When a resident violated discursive norms, others responded with silence, neither engaging nor explicitly criticizing, but allowing the post to hang without response until the conversation moved past it. In November 2024, one resident posted a link to an overseas Chinese-language news article about economic difficulties, adding, "Reporting from outside the wall (note: the Great Firewall, referring to the Chinese state's censorship of global Internet), everyone take a look at what the real situation is." The post received no responses. No one commented. The next message, posted two hours later, was about a neighborhood restaurant's new menu. The poster, a relatively active participant, became noticeably quieter in subsequent weeks.

Moralized critique. When redirection and silence failed, explicit criticism sometimes emerged, framed not in political terms but as character judgment. Those who expressed pessimism or complaint were accused of "传播负能量" ("spreading negative energy"), being "unconstructive," or "only knowing how to complain, not solve problems." This framing transformed political speech into personal failings, subject to moral sanction rather than political repression.

A particularly vivid exchange occurred in April 2024, when a resident complained repeatedly about various aspects of life in the compound. Another resident responded:

"Some people spread negative energy in the group every day, bringing everyone's mood down. If you don't like it here, you can move away; no one's stopping you. Those of us who stay want to live well; we don't need you talking things down every day."

HOA Group 2, April 2024

The response illustrates several mechanisms at once. The moralization of complaint as "negative energy" (for interpretation of the "positive/negative energy" discourse, see Hizi 2021; Yang and Tang 2018). The threat of exclusion. The construction of a collective "we" defined against the complainer. Notably, the original complaints had been about apolitical topics like construction noise and business quality, but the comprehensive pessimism they implied was treated as transgressive. The message was clear. Maintain the appearance of acceptable conditions, or face social sanction.

This peer enforcement explains why state intervention remained minimal throughout the observation period. The grid workers rarely needed to act because the community regulated itself. The occasional reminder of state presence, a policy announcement or a request for cooperation with some inspection, was sufficient to activate the community's

internal discipline. The state established the infrastructural and symbolic conditions under which residents internalized and enacted discipline. Grid workers, surveillance infrastructure, and occasional enforcement created the framework within which the residents themselves operated the machinery.

Punishing Exit

The previous section documented how silence was produced and enforced within the community. But silence about what? Among the most aggressively policed topics was any discourse suggesting that alternatives to the current situation existed. The community's response to exit, both emigration and below-market property sales, reveals a specific and revealing form of discursive control. Exit was not merely an individual choice but a threat to collective meaning-making. Those who exited, or who discussed exit, faced fierce community sanction.

During the darkest weeks of the lockdown, discussion of emigration briefly appeared in community groups. The term " 润 " (rùn), a homophone pun on "run" meaning to leave China, had gone viral on social media, and a few residents shared information about visa categories and overseas property markets. In late April 2022, as the lockdown extended beyond all announced end dates, one resident posted:

"Does anyone know about emigration to the Caribbean countries? Not joking, seriously considering it. This situation has made me see a lot of things clearly… My child is still young. I don't want him growing up in this kind of environment."

Group Purchase Group 3, April 28, 2022

The response was immediate and multilayered, drawing on several registers. Some invoked patriotism: "Thinking of running when the country faces difficulty? This kind of person… forget it, I won't say more." Others framed opposition in practical terms: "As if emigrating is so easy? Language, work, cultural differences… besides, how serious is the pandemic abroad right now, going out to die?" Still others expressed class resentment: "People with money to 'run' can certainly consider it. The rest of us ordinary people should just stay put. What's the point of talking about this?" The original poster attempted a defense but did not raise the topic again. Within days, emigration-related discussion had effectively ended in all groups I monitored. The topic had been established as illegitimate through community sanction, not state censorship.

Yet people did leave. Several residents emigrated during the observation period. I am aware of at least eight families who left, based on property sales, group departures, and neighbor conversations. Rather than being forgotten or wished well, emigrants became negative reference figures whose departure was narrated as betrayal and whose occasional comments were treated as provocation.

The incident described in my introduction, an emigrant mocking the compound's declining property values before being attacked until she left the group, was not isolated. In August 2023, a family that had moved to Singapore posted a photo of their new apartment with the caption "New life begins, slowly adjusting to everything." The response was overwhelmingly hostile.

"You ran away and still post this stuff, showing off?"

"If life abroad is good, then live it quietly. Why stay attached to this group? Want to prove your choice was right?"

HOA Group 1, August 2023

The emigrant's post had been innocuous, a simple life update. But it was received as an implicit critique. Her new life represented a judgment on the lives of those who remained. The accusation of "showing off" transformed her sharing into aggression. The question "want to prove your choice was right?" revealed the underlying anxiety that perhaps it was. The family reduced their group participation and eventually left entirely.

This pattern reveals something important about the informational function of exit in Hirschman's framework. Exit is not merely an individual response to decline. It is also a signal to remaining members about organizational quality. By attacking emigrants and suppressing exit discourse, the community blocked this informational function. What provoked the fiercest sanction was not the act itself, since several families did leave without public incident, but the act of making the exit legible as a rational choice. The emigrant in the opening vignette was not punished for leaving but for narrating her departure as vindication. Low-price sellers were not condemned for selling but for making visible that the market had turned. Departure could not serve as a signal because it was immediately reframed as a moral failing rather than a rational response to conditions.

A parallel dynamic of exit-punishment operated around property sales. As the market declined through 2023–2024, some residents accepted prices below recent benchmarks to facilitate quick sales. These "low-price sellers" became targets of intense community hostility. In February 2024, news circulated that a unit in Building 7 had sold for 62,000 RMB per square meter, approximately 18,000 below the price achieved by a similar unit in 2021. The discussion was furious.

"Selling at this price, trying to drag down the whole compound's property values? Rushing to cash out yourself while making everyone else's assets shrink. So selfish."

HOA Group 2, February 2024

"If everyone holds firm and doesn't sell low, the market will naturally stabilize. It's people like this who panic-sell that keep driving prices down."

Same thread, February 2024

The logic was economically questionable, since individual sellers have minimal impact on market-wide price trends, but emotionally compelling. By blaming departing neighbors for declining values, remaining residents could maintain the fiction that property prices were within community control. Selling became a statement about faith in the city, in the future, in the system that had promised property ownership as the path to get ahead in the wealth race. The phrase " 对上海没有信心 " ("no confidence in Shanghai"), which appeared in the same thread, is particularly revealing. Low-price sellers were not merely pursuing individual interests but expressing a judgment that implicated everyone who remained.

Taken together, the suppression of emigration discourse and the stigmatization of low-price sellers reveal a systematic closure of alternatives. These findings require substantial modification of Hirschman's framework. In my field site, exit exists but is discursively blocked. People do emigrate, but emigration cannot be discussed as a legitimate option without triggering community sanction. The informational function of exit is nullified because leavers are immediately recategorized as traitors whose judgment is suspect. Voice is constrained by the surveillance and peer discipline mechanisms documented in the first section. What remains is not loyalty in Hirschman's sense, which implies belief in the organization's value and potential for reform. The residents I observed rarely expressed such beliefs in group discourse. In informal conversations, many displayed sophisticated cynicism about the system's failures. Yet they

remained, and they enforced norms of silence and staying.

I term this condition "cynical entrapment." Its observable features include discourse patterns suggesting recognition that conditions have declined, interactional evidence that both exit and voice carry unacceptable social costs, remaining not from expressed loyalty but from perceived lack of alternatives, and maintaining this position through collective suppression of information about alternatives. Cynical entrapment is clearly self-reinforcing. Suppressing exit discourse prevents information about alternatives from circulating, making exit seem less viable and further suppressing exit discourse. The system maintains itself by closing off the alternatives that might destabilize it.

Fractured "Community"

If voice is suppressed by peer discipline and exit is discursively blocked, where do grievances go? The third finding suggests an answer. Rather than accumulating into collective pressure against the state, frustration was redirected horizontally, into conflicts among community members themselves. Deep divisions within the ostensibly unified "middle-class community" functioned to absorb grievances that might otherwise have targeted the political system. Beneath the surface of shared status lay profound cleavages that repeatedly erupted in horizontal conflict, consuming energy that might have fueled collective action.

The most visible fault line ran between property owners and renters. Both groups lived in the compound, participated in the same WeChat groups, and shared the same physical space. But their interests diverged on the central question of property values. Homeowners experienced falling values as a mental crisis as their primary asset eroded and their retirement security was threatened. Renters might theoretically benefit from lower prices. When (suspected) renters expressed any hint of this perspective, the response was fierce.

In October 2024, during a discussion of falling prices, a renter observed:

"Actually, housing prices coming down a bit is good for young people. My cousin has been working for five or six years and still can't afford to buy. Now she finally sees some hope."

HOA Group 1, October 2024

The reaction was immediate and harsh:

"You're a renter, right? Of course you think falling prices are good. Those of us carrying millions in mortgage debt, paying more each month than your rent, watching our assets shrink, how are we supposed to see 'hope'?"

"What right does a renter have to comment on property prices? You'll live here a few years and leave; we've bet our life savings on this place."

Same thread, October 2024

The accusation that renters "have no stake" and therefore "no right to comment" demonstrates a conception of community membership as property-based. Only those with financial exposure to property values are legitimate stakeholders. Those without property investment are guests, transients, outsiders, regardless of how long they have lived there. The (suspected) renter who offered the original comment did not respond to the attacks and did not raise similar perspectives again.

A second cleavage, overlapping with but distinct from the owner-renter divide, ran between Shanghai natives (本地人) and migrants (外地人). This division carried cultural and symbolic weight beyond economic interest. The compound's demographics created an ironic situation. As described earlier, the roughly 20 percent who were Shanghai natives had

mostly acquired apartments through demolition compensation without the educational or occupational achievements typically associated with middle-class status. The majority of other owner-occupants were migrants who had come to Shanghai for university or employment, built professional careers, and purchased at market price with substantial mortgage debt. In conventional measures of cultural capital, they far exceeded the demolition-compensation recipients.

Yet in community interactions, the status hierarchy was often inverted. Local residents claimed superior status by virtue of their Shanghai identity and cost-free ownership, sometimes referring to migrants, including university professors, engineers, and doctors, as "外地人" in tones carrying distinct condescension, or worse, "乡下人" ("country people"). A sharp exchange occurred in July 2024.

"The compound's property fee is too high and service doesn't match. Can the homeowners' committee negotiate with property management?"

"You outsiders don't understand Shanghai's market. This price is cheap for Shanghai. If you think it's expensive, you can go back to your hometown. Property fees are definitely lower there."

HOA Group 1, July 2024

The dismissal "you outsiders don't understand" delegitimized the complaint by categorizing the speaker as ignorant. The suggestion to "go back to your hometown" echoed the rhetoric deployed against emigrants and low-price sellers. If you are dissatisfied, leave. The migrant resident responded defensively, "I've lived in Shanghai for fifteen years, bought property here before you did." But the exchange left a residue of hostility. What matters in community standing is not education, occupation, or achievement but origin and ownership. A factory worker whose family received demolition compensation outranks, in community status terms, a physician who purchased at market price and carries mortgage debt.

What made these internal divisions politically significant was their system-maintaining function. By directing frustration toward fellow community members, residents avoided directing it toward the structural conditions and policy choices that actually determined their circumstances. The energy that might have fueled collective grievance was dissipated in horizontal status competition.

Consider the pattern of complaints during the economic downturn. Property values declined because of national real estate policies, demographic trends, the COVID aftermath, and regulatory crackdowns on multiple sectors. These structural causes were entirely beyond community control. Yet discussion in the WeChat groups rarely engaged them. Instead, residents blamed neighbors who sold at low prices, renters who "don't care about the community," migrants who "don't understand Shanghai," the property management company, and specific officials at the most local level. What was systematically absent was any attribution of responsibility to higher levels of government, to national policy, or to the political-economic system that had produced both the property boom and its collapse.

The closest residents came to systemic critique was occasional sardonic humor. In November 2024, when someone asked why property prices kept falling, one resident replied: "Everyone who asks this question has been taken away." The joke got laughing emojis but no substantive responses. It acknowledged the unspeakable, that policy was responsible, that questioning policy was dangerous, while maintaining the pretense that nothing serious had been said. This pattern corresponds to displaced aggression, the redirection of frustration from a threatening target to a safer one. Blaming the central government is dangerous. Blaming

neighbors, renters, migrants, or property management is safe.

These findings challenge the analytical utility of "middle class" as a unified category in the Chinese context. The residents of this compound share a similar economic position despite ownership and might be classified together in any survey-based study. Yet they lack what classical sociology considers essential for class formation. Their interests diverge on property values. Their identities are fractured along lines of origin and ownership. Their capacity for collective action is undermined by these very divisions, which operate along at least three distinct axes, namely tenure-based fracture between owners and renters with divergent material stakes in property values, origin-based fracture between Shanghai natives and migrants carrying symbolic weight beyond economic interest, and acquisition-based fracture between demolition recipients and market purchasers with different relationships to property debt and risk. These cleavages do not merely coexist but cross-cut and reinforce one another, which makes solidarity along any single dimension nearly impossible. This fragmentation helps explain why economic grievances do not produce mobilization. Relative deprivation theory (Gurr 1970) predicts that unmet expectations breed collective action, but the theory requires group identification as a mediating condition. In a community where residents identify against each other rather than with each other, this condition is not met.

The three mechanisms documented above operate in concert to produce what I call a "自律的中产阶级" (self-disciplining middle class). This is not a class beaten into submission by state repression, nor one that genuinely believes in the system's legitimacy. It is a class that actively participates in maintaining order through its own internal dynamics. Community members monitor each other's speech, enforcing boundaries that state actors rarely need to police directly. They punish deviations through social sanction rather than legal punishment. They suppress information about alternatives. And they redirect their grievances horizontally by attacking fellow residents rather than the structures that constrain them all.

This configuration means that the Chinese state maintains control not only through continuous direct repression but through establishing the conditions under which communities regulate themselves. The state need not monitor every WeChat group or censor every critical comment when residents enforce discursive boundaries on its behalf. The community does much of this work itself. It means that economic decline does not automatically generate political challenge, because grievances can be absorbed by internal divisions and displaced onto safe targets. Sloterdijk's (1987) concept of "enlightened false consciousness" captures the subjective dimension. The modern cynic knows what they are doing but does it anyway, not from naïveté but from resignation and recognition that alternatives are unavailable. The residents I observed were not naive. Many made sardonic jokes revealing sophisticated awareness. Yet this awareness produced not resistance but compliance, sustained through collective enforcement of discursive norms that rendered alternatives unspeakable.

Conclusion and Discussion

This paper examines the political attitudes and behaviors of China's urban middle class during and after the 2022 Shanghai lockdown. I documented three mechanisms by which middle-class communities maintain political quiescence. Surveillance-induced self-censorship and peer discipline produce a progressive deepening of silence. The punishment of exit and discursive closure of alternatives block the informational function of departure. And the

displacement of grievances onto horizontal conflicts between homeowners and renters, locals and migrants, absorbs energy that might otherwise fuel collective challenge. Together, these mechanisms constitute a self-disciplining middle class that actively participates in maintaining social order through its own internal dynamics.

The first contribution extends the critique of modernization theory. Existing scholarship has established that China's middle class does not conform to the expectation that economic development generates demands for political liberalization (Chen 2013; Dickson 2016; Wright 2010). This study goes further by showing that the Chinese middle class not only fails to demand liberalization but actively participates in suppressing those who might. The obstacle to middle-class politics in China is not merely state repression or co-optation but the internal structure of the class itself. Its fragmentation along property and identity lines, its property-based definition of membership, and its mechanisms for enforcing conformity all work to prevent collective action. The classical expectation from Lipset through Inglehart and Welzel assumed that the middle class would develop shared interests and common identity as a natural consequence of lifted economic position. The Chinese case reveals that "middle class" can designate a property-defined stratum where residents share income levels and consumption patterns but have divergent interests, fractured identities, and no capacity for unified action. Property is not a foundation for political agency but an anchor preventing its exercise.

The second contribution complicates relative deprivation theory. Shanghai's middle class experienced precisely the gap between expectations and outcomes that Gurr(1970) identified as the wellspring of collective action. The lockdown violated the implicit "trade freedom for material improvement" social contract in reform-era China. The economic decline contradicted expectations of continued prosperity. Yet mobilization did not follow. Two mechanisms blocked the translation of grievance into action. Internal fragmentation prevented the formation of "fraternalistic" relative deprivation (Walker and Smith 2002), the perception that one's group is disadvantaged. When residents identify against each other rather than with each other, grievance remains individual rather than collective. Meanwhile, the systematic suppression of exit discourse blocked what might be called exit-induced voice, the possibility that seeing others leave might prompt remaining members to reconsider their situation. When emigrants are recategorized as traitors and low-price sellers as defeatists, their departure cannot serve as information. It becomes instead evidence for the collective fiction that staying is the only rational choice.

The third contribution revises Hirschman's (1970) Exit-Voice-Loyalty framework for contemporary techno-authoritarian contexts. I propose the concept of "cynical entrapment" to describe a condition where exit is blocked discursively through community sanction even when it remains practically available, voice is constrained by surveillance and peer discipline, and what remains is neither loyalty nor any viable alternative. The key difference from Hirschman's loyalty is the apparent absence of affirmative commitment. Loyal members believe things can improve and work toward that improvement. The cynically entrapped display patterns of resigned compliance suggesting they doubt improvement is possible yet see no way out. This distinction matters because it suggests different dynamics. Loyalty can be won or lost through organizational performance. Cynical entrapment is maintained by closing off alternatives regardless of performance. The system need not respond to its members because its members will remain regardless.

Altogether, my findings shift analytical attention from vertical state-society relations

to the horizontal relations within society that make vertical control possible. The dominant paradigm in studies of Chinese authoritarianism emphasizes state capacity, including sophisticated censorship (King, Pan, and Roberts 2013; Roberts 2018), adaptive governance (Nathan 2003), and gridded surveillance (Mittelstaedt 2022). This study does not challenge the importance of state capacity but suggests it tells only part of the story. The state establishes the parameters within which social control operates, through the presence of grid workers, through occasional enforcement, and through the infrastructure of surveillance. But communities do much of the daily work of maintaining order themselves. This state-enabled, socially enacted control is both more pervasive and less costly than direct repression. Once established, it operates continuously without requiring constant state resources, and it carries no legitimacy costs because the discipline appears to originate from society itself.

My argument that internal fragmentation absorbs potential resistance adds another dimension. The state need not actively divide and rule because divisions emerge organically from the structure of property-based class formation. When middle-class identity is defined by property ownership, those with different property situations develop divergent interests. When local registration carries material and symbolic weight, locals and migrants become competitors rather than allies. This analysis suggests that economic decline alone will not generate political challenges to authoritarian rule. The translation of grievance into collective action is blocked at multiple points, including the absence of a shared identity, the suppression of information about alternatives, peer enforcement of conformity, and the displacement of frustration onto safe targets.

Yet the foundation of middle-class acceptance, confidence that the system delivers prosperity and security, has been substantially eroded. The residents I observed understood this. Their cynicism was not naïve. They recognized that the promises underlying their compliance had been violated. The sardonic jokes, the careful circumlocutions, and the sophisticated evasions all revealed awareness that could not be openly expressed. Whether this withdrawal represents a stable equilibrium or a transitional state remains a pending question. The mechanisms documented here have maintained quiescence through three years of sustained difficulty. But cynical entrapment is an equilibrium maintained by the absence of perceived alternatives, not by satisfaction.

Admittedly, several limitations constrain these findings. The study is based on a single compound in one district of Shanghai, a city that is significantly wealthier, more cosmopolitan, and was more heavily affected by the 2022 lockdown. The mechanisms I document may operate differently in other contexts. The data source, WeChat group discussions, captures semi-public discourse but not private sentiment. The silence I document is silence in community forums. Residents may express very different views in trusted private conversations, and the relationship between performed compliance and private belief is a persistent challenge in research on authoritarian societies (Scott 1990). My position as a non-local renter at the bottom of the community's status hierarchy shapes both my access and my interpretation. A Shanghai native homeowner might observe different dynamics or have access to discussions from which I was excluded. Finally, the observational methodology cannot establish causation. I have documented patterns and proposed mechanisms to explain them, but alternative interpretations cannot be ruled out.

Future research could address these limitations through comparative studies across different types of communities, research on emigrants who have exercised the exit option, and longitudinal observation extending beyond the current period. Combining digital ethnography with survey methods could assess the relationship between public silence and private sentiment, while analysis of communities that have experienced successful collective action could identify what conditions enable mobilization to overcome the barriers documented here.

In the end, I want to go back to the resident's joke about property prices ("everyone who asks this question has been taken away"). It uniquely captured something essential about the condition I have documented. The joke acknowledged what could not be openly stated, and it revealed sophisticated awareness behind performed compliance. In addition, it demonstrated the characteristic mode of expression under cynical entrapment, speech that communicates through indirection, that everyone understands and no one dares to acknowledge. Therefore, this is not the silence of the ignorant or the cowed.

The implications extend beyond China. As digital platforms increasingly mediate community life, as property becomes central to middle-class identity globally, and as authoritarian governance adapts to contemporary conditions, the mechanisms documented here may have wider relevance. Self-disciplining communities, cynical entrapment, and the displacement of vertical grievance into horizontal conflict are not uniquely Chinese phenomena but responses to structural conditions that exist, in varying degrees, in many societies.

For China specifically, this study suggests that the stability of authoritarian rule rests on foundations both more robust and more fragile than commonly assumed. More robust because control is distributed through social structures rather than concentrated in state institutions. More fragile because it depends on the continuing closure of alternatives. The middle class that disciplines itself today might, under different conditions, become the agent of transformation it has so far failed to be. After all, as Qin Hui (2004) persuasively argues, even if something like "national character" (民族性) exists, it cannot serve as the basis for any form of historical determinism.

Acknowledgments

I thank the reviewers and the editors at JLMS for their comments on earlier drafts.

This article was completed on March 28, 2026, four years to the day after Shanghai entered lockdown in 2022. What began as confinement became, unexpectedly, the origin of this line of research. The Shanghai lockdown was, for those who lived through it, a rupture. For me it also broke open a set of questions I had not known how to ask. I dedicate this work to everyone who endured that spring and its long aftermath.

I owe so much to the residents of the Baoshan compound who, without knowing they were being studied, showed me genuine contemporary Chinese lives and thoughts. Their words, emotions, and quarrels form the substance of this paper.

Disclosure Statement

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

Funding Information

This research received no external funding.

Data Availability Statement

Research data are not to be shared for 1) they contain original and identifiable information of human subjects; 2) they potentially contain politically sensitive information that may trigger risks for the individuals documented in this research.

References

Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." Qualitative Research in Psychology 3 (2): 77–101.

Braun, Virginia, and Victoria Clarke. 2022. Thematic Analysis: A Practical Guide. London: SAGE.

Cai, Yongshun. 2005. "China's Moderate Middle Class: The Case of Homeowners' Resistance." Asian Survey 45 (5): 777–99.

Chau, Grace W. F., and Sergiu Gherghina. 2024. "Conditional Loyalty and Exit: Explaining the Emigration of Wealthy Chinese after the 2012 Leadership Change." Diaspora Studies 17 (2): 113–133.

Chen, Jie. 2013. A Middle Class Without Democracy: Economic Growth and the Prospects for Democratization in China. Oxford: Oxford University Press.

Clark, William Roberts, Matt Golder, and Sona N. Golder. 2017. Principles of Comparative Politics. 3rd ed. Washington, DC: CQ Press.

Davis, Deborah S. 2003. "From Welfare Benefit to Capitalized Asset: The Re-commodification of Residential Space in Urban China." In Housing and Social Change: East-West Perspectives, edited by Ray Forrest and James Lee, 183–98. London: Routledge.

Dickson, Bruce J. 2016. The Dictator's Dilemma: The Chinese Communist Party's Strategy for Survival. New York: Oxford University Press.

Dwyer, Sonya Corbin, and Jennifer L. Buckle. 2009. "The Space Between: On Being an Insider-Outsider in Qualitative Research." International Journal of Qualitative Methods 8 (1): 54–63.

Foucault, Michel. 1977. Discipline and Punish: The Birth of the Prison. Translated by Alan Sheridan. New York: Pantheon Books.

Franzke, Aline Shakti, Anja Bechmann, Michael Zimmer, Charles Ess, and the Association of Internet Researchers. 2020. Internet Research: Ethical Guidelines 3.0. Association of Internet Researchers. https://aoir.org/reports/ethics3.pdf.

Gurr, Ted Robert. 1970. Why Men Rebel. Princeton, NJ: Princeton University Press.

Hine, Christine. 2015. Ethnography for the Internet: Embedded, Embodied and Everyday. London: Bloomsbury Academic.

Hirschman, Albert O. 1970. Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States. Cambridge, MA: Harvard University Press.

Hizi, Gil. 2021. "Zheng Nengliang and Pedagogies of Affect in Contemporary China." *Social Analysis* 65 (1): 23–43. https://doi.org/10.3167/sa.2020.650102

Huang, Haifeng. 2015. "Propaganda as Signaling." Comparative Politics 47 (4): 419–44.

Inglehart, Ronald, and Christian Welzel. 2005. Modernization, Cultural Change, and Democracy: The Human Development Sequence. Cambridge: Cambridge University Press.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." American Political Science Review 107 (2): 326–43.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument." American Political Science Review 111 (3): 484–501.

Kozinets, Robert V. 2019. Netnography: The Essential Guide to Qualitative Social Media Research. 3rd ed. London: SAGE.

Lipset, Seymour Martin. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." American Political Science Review 53 (1): 69–105.

Mittelstaedt, Jean Christopher. 2022. "The Grid Management System in Contemporary China: Grass-Roots Governance in Social Surveillance and Service Provision." China Information 36 (1): 3–22.

Moffa, Grazia, and Marco Di Gregorio. 2023. "Exploring the Use of WeChat for Qualitative Social Research: The Case of Italian Digital Diaspora in Shanghai." Frontiers in Sociology 8: 1144507.

Moore, Barrington, Jr. 1966. Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World. Boston: Beacon Press.

Nathan, Andrew J. 2003. "China's Changing of the Guard: Authoritarian Resilience." Journal of Democracy 14 (1): 6–17.

O'Brien, Kevin J., and Lianjiang Li. 2006. Rightful Resistance in Rural China. Cambridge: Cambridge University Press.

Ong, Lynette H. 2023. "China's Epidemic of Mistrust: How Xi Jinping's COVID-19 U-Turn Will Make the Country Harder to Govern." Foreign Affairs, January 11, 2023.

Perry, Elizabeth J. 2008. "Chinese Conceptions of 'Rights': From Mencius to Mao—and Now." Perspectives on Politics 6 (1): 37–50.

Przeworski, Adam, Michael E. Alvarez, José Antonio Cheibub, and Fernando Limongi. 2000. Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990. Cambridge: Cambridge University Press.

Qin, Hui. 2004. *Shijian ziyou* [Practising Freedom]. Hangzhou: Zhejiang People's Publishing House.

Roberts, Margaret E. 2018. Censored: Distraction and Diversion Inside China's Great Firewall. Princeton, NJ: Princeton University Press.

Robinson, Darrel, and Marcus Tannenberg. 2019. "Self-Censorship of Regime Support in Authoritarian States: Evidence from List Experiments in China." Research & Politics 6 (3): 1–9.

Rueschemeyer, Dietrich, Evelyne Huber Stephens, and John D. Stephens. 1992. Capitalist Development and Democracy. Chicago: University of Chicago Press.

Scott, James C. 1990. Domination and the Arts of Resistance: Hidden Transcripts. New Haven,

CT: Yale University Press.

Sloterdijk, Peter. 1987. Critique of Cynical Reason. Translated by Michael Eldred. Minneapolis: University of Minnesota Press.

Tomba, Luigi. 2014. The Government Next Door: Neighborhood Politics in Urban China. Ithaca, NY: Cornell University Press.

Walker, Iain, and Heather J. Smith, eds. 2002. Relative Deprivation: Specification, Development, and Integration. Cambridge: Cambridge University Press.

Wang, Jinpu, and Yu Xiang. 2025. "Plague, Food, and Freedom: Bounded Resistance in the Shanghai COVID-19 Lockdown." Sociological Focus 58 (4): 471–500.

Wright, Teresa. 2010. Accepting Authoritarianism: State-Society Relations in China's Reform Era. Stanford, CA: Stanford University Press.

Xie, Yu, and Yongai Jin. 2015. "Household Wealth in China." Chinese Sociological Review 47 (3): 203–29.

Yang, Tony Zirui. 2025. "Normalization of Censorship: Evidence from China." The Journal of Politics 87 (4): 1227–42.

Yang, Peidong, and Lijun Tang. 2018. "'Positive Energy': Hegemonic Intervention and Online Media Discourse in China's Xi Jinping Era." *China: An International Journal* 16 (1): 1–22. https://doi.org/10.1353/chn.2018.0000.

Jinpu Wang is an assistant professor of sociology and anthropology at Metropolitan State University (Minnesota, USA). His research examines multiple streams of Chinese emigration to the West and Africa triggered by political-economic changes in contemporary China, as well as digital authoritarianism and contentious politics within China. He holds a Ph.D. in Sociology from Syracuse University. Correspondence should be addressed to the author at jinpu.wang@metrostate.edu.

Journal of Language, Media and Society
Vol. 1, No. 1 https://doi.org/10.66056/jlms.2026.v1i1.006

Carter & Co. PUBLISHING

FROM FAIRY TALES TO YOUNG ADULT: A REVIEW OF *THE ROUTLEDGE HANDBOOK OF TRANSLATION AND YOUNG AUDIENCES*

LIJUAN XU & JUAN ZHANG

This review evaluates *The Routledge Handbook of Translation and Young Audiences* (2025), co-edited by Michal Borodo and Jorge Díaz-Cintas, a landmark volume that formally establishes Translation for Young Audiences (TYA) as an independent discipline. The authors analyze the handbook's contributions across theoretical, methodological, and practical dimensions. Its primary significance lies in driving a paradigm shift from text-centric to audience-centered approaches, introducing the "agentic reader" and "double dialogue" models. By integrating corpus stylistics, neurocognitive eye-tracking, and multimodal analysis, the handbook propels TYA into a rigorous empirical stage. Ultimately, this work is of profound importance for redefining translation as a creative, intergenerational cultural practice within contemporary media and society.

KEYWORDS: TYA; Paradigm Shift; Interdisciplinary; Multimodality; Agentic Reader

Introduction

Against the backdrop of digital transformation, the cultural consumption patterns of young audiences have undergone a radical metamorphosis. Translation, as a vital mechanism for cross-cultural mediation, must now adapt to these shifting demographics and media environments. *The Routledge Handbook of Translation and Young Audiences* (2025), co-edited by Michal Borodo and Jorge Díaz-Cintas, synthesizes the contributions of 47 scholars from 25 countries. The volume transcends traditional literary translation boundaries, advocating for a "young audience-centered" research agenda. Its publication serves as a formal landmark for the maturation of TYA as an independent academic field.

Content Overview

The *Handbook* constructs a comprehensive knowledge system organized into five sections and 35 chapters, bridging theory, text, medium, and practice.

Part I (Ch. 1–8) establishes theoretical foundations. It innovatively proposes the "agentic reader" (Ch. 1), breaking the paternalistic view of children as passive recipients. Chapter 2 utilizes Bakhtin's dialogism to build a "double dialogue" model, balancing the translator's fidelity to the author with their responsibility toward the child's cognitive expectations. Chapter 7 introduces neurocognitive data via eye-tracking experiments, providing empirical evidence for how children process translated content.

© 2026 Carter & Co.

Part II (Ch. 9–14) re-examines literary practices. It delves into archival research on ideological interventions in 20th-century Poland and analyzes Indigenous adaptations, such as the Australian Pitjantjatjara version of *Alice in Wonderland*, which utilizes cultural relocation strategies (e.g., replacing the White Rabbit with a kangaroo).

Part III (Ch. 15–21) focuses on Audiovisual Translation (AVT). It contrasts the localization of *Frozen* across 12 languages and explores the "dynamic graded subtitling system" (Ch. 18), which significantly enhances comprehension for D/deaf children by adjusting reading speeds (80–120 words/minute).

Part IV (Ch. 22–29) explores emerging media, including word-image intertextuality in picture books and the complex localization workflows of video games.

Part V (Ch. 30–35) envisions the digital future, discussing participatory translation in *Genshin Impact*, Harry Potter fan communities in China, and the ethical frameworks of "co-creation" where children participate in translation decision-making.

Critical Evaluation

As a significant milestone in the field of translation studies, *The Routledge Handbook of Translation and Young Audiences* demonstrates groundbreaking scholarly value across three dimensions: theoretical construction, methodological innovation, and practical application. The volume systematically fills a long-standing lacuna in research on translation for children and adolescents (TYA) and promotes a fundamental paradigm expansion from a traditional, text-centric focus to an interdisciplinary, audience-oriented perspective. The following sections evaluate the handbook's contributions within these three dimensions and discuss the future trajectories it establishes for the field.

Theoretical Evolution: The Triadic Model

The handbook shatters the adult-centric bias that views children's translation as a "simplified form" of adult work. Drawing on Oittinen (2000), it asserts that "situation and purpose are inherent in all translation". This theoretical breakthrough integrates Hermans' (2019) framework of translation ethics, emphasizing that translators must assume ethical responsibility for the child reader's cognitive level by adopting a "child-cognitive-development-oriented" strategy. This ethical turn signifies a shift from treating children as "miniature adults" to recognizing their distinct audience needs.

By integrating Bakhtin's dialogism, the volume proposes a "double dialogue" model, where translators balance the artistic dialogue with the author and the anticipatory interaction with the child reader. This reflects a transition from a binary "Text–Translator" structure to a triadic "Text–Reader–Medium" structure (Bassnett, 2014), highlighting the impact of multimodal elements on children's meaning construction.

Furthermore, the application of "Third Space" theory (Bhabha, 1994) highlights children's cultural hybridity. This is exemplified in the Pitjantjatjara Indigenous translation of Alice in Wonderland, where the White Rabbit is replaced by a kangaroo to align with the child's cognitive framework (Malmkjær). The handbook critically expands functionalism, advocating for a "uniquely adaptive approach" (Oittinen, 2000) rather than mechanical equivalence, as seen in Danish Winnie-the-Pooh translations (Malmkjær, 2025) that prioritize narrative expectations over formal fidelity.

Methodological Rigor: The Empirical Turn

The volume pushes TYA toward experimental science. Corpus stylistics (Malmkjær, 2025) quantifies the "readability first" principle, showing a 37% reduction in sentence length and 52% drop in compound sentences in Danish translations. Cognitive science tools like eye-tracking (Lozano, 2025) provide neurocognitive proof that cultural adaptation reduces fixation duration by 19% and increases comprehension accuracy by 23% for children aged 6–8.

Additionally, the multimodal analysis model (Kaindl, 2025; Zanettin, 2014) and ethnographic approaches to fan translation (Chan, 2025; Venuti, 2008) offer a robust toolkit for analyzing how images, typography, and online collaboration reconstruct the translator's role in the digital age.

Practical Innovation: Accessibility and Participation

In practice, the handbook demonstrates that cultural adaptation is a spectrum based on audience cognition (Minutella, 2025). Its humanistic contribution is evident in Zárate's (2025) dynamic graded subtitling, which improved comprehension for d/Deaf children by 55%. Regarding AI empowerment, Fu (2025) and the handbook both emphasize a human-AI collaboration where AI handles linguistic simplification while human translators ensure affective and cultural fidelity. Finally, it highlights inclusive practices like "multisensory translations" for visually impaired children, embodying social inclusion.

Limitations and Future Prospects

Despite its achievements, the handbook's limitations offer insights into future research. First, a regional imbalance persists. As Peng (2024) notes, the demand for localized theoretical shaping in China creates a tension with the handbook's Western-centric focus. While Hou (2025) illustrates that social systems and translation are mutually constitutive, the volume overlooks critical contexts like China's "Double Reduction" policy or Arab religious adaptations. Since translation is a dynamic "knowledge reconstruction" rather than static transfer (Wu, 2019), future studies should employ semiotic methodology to examine how diverse child readers act as "interpretants" (Wang, 2019). This aligns with the "Sublimation (Huajing)" theory, which Li (2025) describes as an open, evolving system—a perspective essential for localized TYA iterations.

Second, technological ethics require deeper engagement. The handbook under-addresses algorithmic bias and data privacy—critical issues for minors in the AI era. Future work must establish ethical frameworks to evaluate AI's impact on children's decoding processes. Finally, theoretical integration remains fragmented. Strengthening the "theoretical grafting" between semiotics and translation will help build a more cohesive model of translation as knowledge reconstruction (Wu, 2019), facilitating the organic generation of non-Western paradigms in this complex, transmedia field.

Conclusion

The handbook's primary value lies in its "intergenerational dialogue" model, positioning children as active co-creators. This elevates translation beyond linguistic conversion into a creative ethical practice (Liu & Xu, 2022) of cultural inheritance. By constructing a "Theory–Medium–Culture–Technology" coordinate system, the volume remains an essential compass for understanding the future of cross-cultural communication.

References

Bassnett, S. (2014). *Translation studies*. Routledge.

Bhabha, H. K. (1994). *The location of culture*. Routledge.

Chan, L. T. (2025). Young readers, young translators: Harry Potter fan translation projects in China [In Chinese]. In M. Borodo & J. Díaz-Cintas (Eds.), *The Routledge Handbook of translation and young audiences.* Routledge.

Fu, J. (2025). New perceptions on criticism standards for AI translation [In Chinese]. *Shanghai Journal of Translators,* (5), 1–7.

Hermans, T. (2019). *Translation in systems: Descriptive and systemic approaches explained.* Routledge.

Hou, J. (2025). System and translation: Translation and cultural reconstruction in The Eastern Miscellany (1904–1948) [In Chinese]. *Shanghai Journal of Translators,* (6), 82–86.

Kaindl, K. (2025). Theoretical frameworks and concepts for studying comics for younger audiences. In M. Borodo & J. Díaz-Cintas (Eds.), *The Routledge Handbook of translation and young audiences.* Routledge.

Li, J. (2025). Important contributions of the first Chinese translation of the English poem "A Psalm of Life" to the "Sublimation (Huajing) Theory" [In Chinese]. *Foreign Languages Research,* (6), 84.

Liu, Y., & Xu, J. (2017). On the positioning of translation and the grasp of translation value: A dialogue [In Chinese]. *Chinese Translators Journal*, (6), 54–61.

Lozano, J. (2025). Eye-tracking as a tool for investigating children's processing of translated texts. In M. Borodo & J. Díaz-Cintas (Eds.), *The Routledge Handbook of translation and young audiences.* Routledge.

Malmkjær, K. (2025). Investigating readability in children's literature translations. In M. Borodo & J. Díaz-Cintas (Eds.), *The Routledge Handbook of translation and young audiences.* Routledge.

Minutella, V. (2025). Translating songs in animated films: A case study of Frozen. In M. Borodo & J. Díaz-Cintas (Eds.), *The Routledge Handbook of translation and young audiences (pp. 201–211).* Routledge.

Oittinen, R. (2000). *Translating for children.* Garland Publishing.

Peng, B. (2024). Centenary history, practical shaping, and theoretical naming: *A review of Research on the History of Chinese Science Fiction Translation over a Century* [In Chinese]. *Chinese Translators Journal*, (6), 98–103.

Venuti, L. (2008). *The translator's invisibility: A history of translation.* Routledge.

Wang, H. (2019). Children's literature translation from the perspective of semiotics and multimodality [In Chinese]. *Chinese Translators Journal*, (3), 124–129.

Wu, S. (2019). On semantic logic in the process of translation: A case study of *The Golden Touch* [In Chinese]. *Chinese Translators Journal*, (5), 152–159.

Zanettin, F. (2014). Visual adaptation in translated comics. *Intralinea*, 1–34.

Zárate, S. (2025). Subtitling for d/Deaf children. In M. Borodo & J. Díaz-Cintas (Eds.), *The Routledge Handbook of translation and young audiences.* Routledge.

Lijuan Xu is a postgraduate student at Huazhong Agricultural University, specializing in audiovisual translation.

Juan Zhang, PhD, is an Associate Professor at the School of Foreign Languages, Huazhong Agricultural University. Her research focuses on audiovisual translation and cultural mediation. She holds a joint PhD from Central China Normal University and UCL Centre for Translation Studies. A visiting scholar at UCLA, Zhang is active in the Translators Association of China and Hubei Translators Association. She has published many papers in top journals and serves as a reviewer for various international publications. Zhang has led multiple research projects, including National Social Science Fund and Hubei Provincial grants. She has received awards for outstanding papers and frequently gives keynote speeches at academic conferences.

This work was supported by the National Social Science Fund of China (Grant No. 24FYYB016) and the Fundamental Research Funds for the Central Universities (Grant No. 2662025WGPY001).