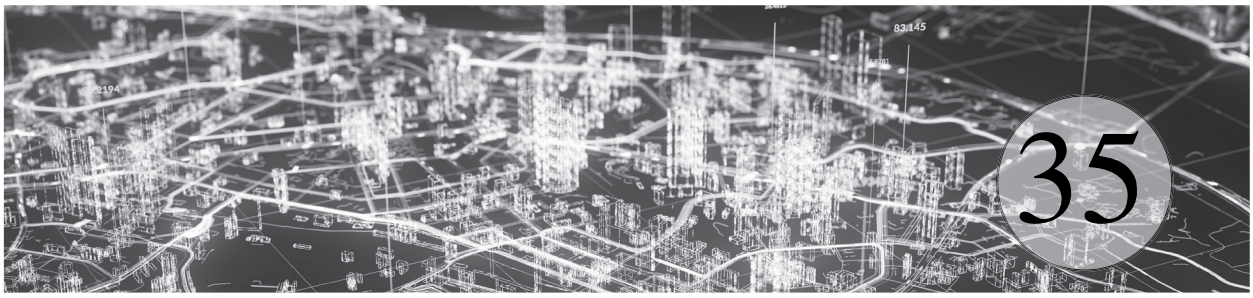

How to Use Computational Methods and Online Experiments to Study Chinese Digital Media

Yingdan Lu and Matt DeButts

INTRODUCTION

China's media ecosystem is among the world's most vibrant and sophisticated. Its 1.079 billion netizens, 1.076 billion mobile Internet users, 3.83 million websites, and 2.6 million active apps (CNNIC, 2023) make it a fertile ground for communication scholars to understand complex and unique communication activities. Its richness and diversity present an excellent opportunity to test new theoretical questions based on novel data, enabling the development of new methodologies in communication research.

This chapter focuses on two pioneering methods—computational methods and online experiments—and provides a systematic reflection of how these methods apply to Chinese communication research. In the years since Lazer et al. (2009) published their groundbreaking delimitation of computational social science as a field, computational methods have spread rapidly across a multitude of social science disciplines and subdisciplines. In communication, researchers such as Shah et al. (2015), van Atteveldt & Peng (2018), van Atteveldt et al. (2019), and Hilbert et al. (2019) have helped to establish computational communication science as a standalone subfield worthy of independent governance and development. These scholars have not only developed the theoretical foundations of computational communication but also highlighted the practical challenges and opportunities that arise when applying computational methods to the multifarious and complex human behaviors that occur on digital media platforms. Concurrently, the internet has evolved into an increasingly attractive option for online

experiments, such as virtual lab experiments and online field experiments. Compared with traditional in-lab and offline experiments, online experiments enable researchers to scale up participant recruitment, better observe group-based dynamics, and strengthen the external validity of research findings (Hilbert et al., 2019).

While computational methods and online experiments are used ever-more-frequently in Western research contexts, there exists a noticeable gap in systematically examining their application within the Chinese media landscape. The distinctive characteristics of China's digital media ecosystem, such as its large volumes of data and the diverse range of activities occurring there, necessitate a deeper understanding of the challenges and opportunities presented for computational methods and online experiments. While prior research has taken an empirical approach to compare computational communication research between Chinese-language and English-language scholarship (Zhao, 2023; Zhong et al., 2023), these studies primarily contrast journal publications rather than focusing on empirical research related to computational communication within the Chinese digital media context. Experimental methods have received even less attention. Research such as Y. Chen et al. (2023) and K. Chen et al. (2024) reviews digital trace research in Chinese communication. However, there is a lack of systematic understanding of computational and experimental methods in studying Chinese digital media. Overlooking these challenges may lead to biased and inaccurate understandings of communication processes, or in the case of overlooked opportunities, a fragmentary view of human communicative behavior.

In light of these considerations, this chapter seeks to offer a comprehensive overview of computational research and online experiments in China, beginning with an empirical analysis of how these techniques have been employed within the study of Chinese communication between 2000 to 2020 by looking at all published work in communication during that time. In the following section, we aim to advance these methods by outlining their specific challenges and opportunities. Our goal is to provide insights for future research endeavors and reflexive considerations for scholars hoping to use these methods to unpack the dynamic Chinese digital media landscape.

This chapter is organized as follows: Section "Empirical-Based Reflections of Using Computational and Online Experimental Methods to Understand Chinese Digital Media" provides an empirical reflection of how computational and online experimental methods have been applied in Chinese communication research from 2000 to 2020. "Rising Interest in Computational Research; Lacking Attention in Online Experiments" critically examines the methodological challenges faced in adapting these methods to the specificities of Chinese digital media. "Methodological Challenges: Rethinking Commensurability, Validity, and Reliability" delineates the unique opportunities that the Chinese digital media landscape offers for computational and experimental research.

"Opportunities: Data Abundance, Dynamic Online Activities, and Theoretical Opportunities" concludes.

EMPIRICAL-BASED REFLECTIONS OF USING COMPUTATIONAL AND ONLINE EXPERIMENTAL METHODS TO UNDERSTAND CHINESE DIGITAL MEDIA

Data and Methods for Empirical-Based Reflections

To acquire an empirical understanding of existing research within the field of Chinese communication research using computational methods and online experimental methods, we collected empirical research papers from 95 communication journals indexed in the Web of Science spanning the years 2000 to 2020. This dataset, previously collected and utilized by Chen et al. (2024), served as the foundation dataset for our analysis. Although "online experiments" are sometimes considered a subset of growing computational methods in prior field-reflection papers (e.g., Hilbert et al., 2019; van Atteveldt & Peng, 2018), we treat online experiments and computational methods as two distinct fields of inquiry, as some authors do not label online experiments as computational studies. We thus extracted targeted article abstracts through querying two sets of keywords: (1) those indicating computational studies in Chinese communication, such as "machine learning" AND "China"; and (2) those indicating experimental studies in Chinese communication, like "random assignment" AND "China" (see Appendix A for complete keyword lists).

After subsetting to abstracts containing these keywords, we further refined our list via manual review. To assess whether the paper contained computational methods, the authors and two trained research assistants manually assessed all abstracts for the use of computational methods as defined by Hilbert et al. (2019). A majority vote was applied in cases of discrepancies. To identify papers using online experiments, our definition of online experiments was drawn from prior definitions in communication research (Hilbert et al., 2019; Muise & Pan, 2019). As the application of online experiments is less visible in abstracts from our pre-classification reading, we used a two-step process. First, two independent coders read each abstract and, when possible, classified the paper as having an online experiment (or not). In many cases, the abstract was not clear whether the experiment occurred online. In those ambiguous cases, we also read the paper's methods section to ensure applicability. In total, we identified 78 computational research papers and 12 research papers using online experiments.

To obtain a better sense of these computational and online experimental works, we classified each paper's major communication subfield with a coding scheme from Chen et al. (2024). We also derived qualitative findings from reading the abstracts for specific topics. For computational work, we also looked into each paper's digital data type and the concrete methods applied in each work.

Rising Interest in Computational Research; Lacking Attention in Online Experiments

As Figure 35.1 shows, the past 20 years have witnessed a growing interest in using computational methods to understand Chinese digital media and communication. The earliest papers using computational research to study China emerged in 2002, with volume increasing steadily over the first decade. In the second decade research accelerated in two bursts: once from 2013 to 2014 (2–7 papers per year), and again between 2018 and 2020 (17 articles).

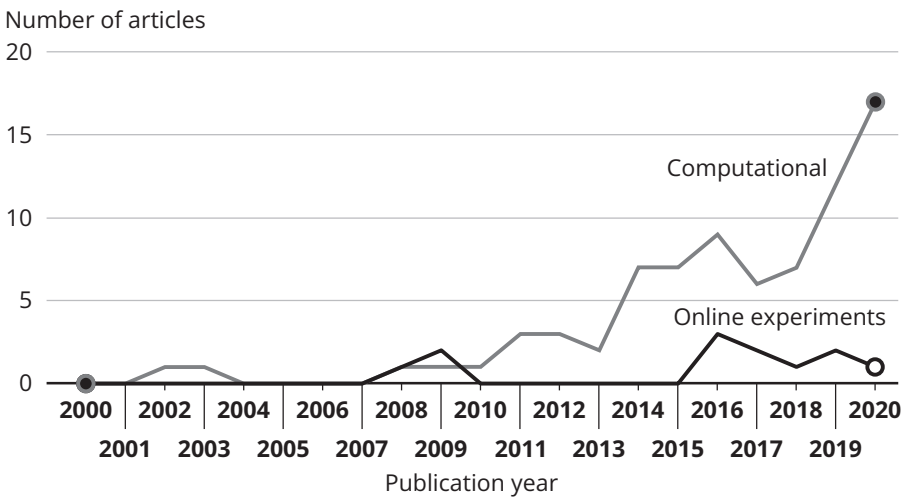

Figure 35.1 Time trend of China-related communication research using computational methods vs. online experiments

Compared with the rapid growth of computational research methods, articles applying online experiments remain low, never exceeding five in a given year. Although the growing penetration of the Internet into everyday life has presumably created more opportunities for virtual experiments, the use of such methods to study Chinese communication remains scarce, potentially as a reflection of the challenges of employing methods in a Chinese communication context (see "Methodological Challenges" section).

Domain and Topics

From the human coding results, we find that researchers predominantly (46%) utilize computational methods for research in political communication. Political protests in mainland China and Hong Kong are a particular focus. For example, Huang and Sun (2016) focus on social media discourse surrounding an anti-nuclear protest in China, while Veneti et al. (2016) explore the similarities and differences in media coverage of the 2014 Hong Kong protests. Studies have also

focused on variation in state internet controls to understand public–government relations, including domestic governance on public opinion (e.g., Lu & Pan, 2021) and external propaganda strategies (e.g., Huang & Wang, 2019). In other cases, political communication topics popular in Western contexts, such as fake news, polarization, and online civility, are also undergoing increased scrutiny in Chinese communication via the use of computational methods.

Computational research has also attracted a notable degree of research interest in the fields of public relations, advertising, and strategic communication, which together account for 14% of all studies in our population. Interestingly, a significant portion of these studies focus on public relations or public diplomacy, i.e. communication phenomena occurring on a global scale (e.g., Wu & Yang, 2017), rather than on advertising, public relations, or strategic communication, i.e. phenomena occurring at the enterprise or organizational levels (e.g., Lu & Li, 2015). Additionally, around 12% of computational studies examine cultural communication in China or between China and other countries, exploring research questions such as national differences in online film ratings (Feng, 2019) and the impact of cultural intelligence on global virtual collaboration (Li et al., 2017).

Unlike studies using computational methods, Chinese communication research using online experiments is not dominated by any communication subfields. The 12 studies using online experiments span multiple communication subfields, including political communication, health communication, environmental communication, media psychology, strategic communication, organization communication, and journalism. There are slightly more studies in health communication, focusing on smoking control (e.g., Bresnahan & Zhuang, 2016) and women's health (e.g., Chen & Yang, 2018).

Data Types

Digital trace data is used far more widely than other data types like traditional news outlets or broadcast transcripts. Within digital trace data, social media data is most common: a huge body of work examines activities occurring on China's largest media platforms, such as Weibo and WeChat, or Chinese government outreach to foreign citizens on international platforms like Facebook and Twitter. Researchers also frequently examine trace data collected from online websites and portals. For example, Huang and Sun (2016) constructed a hyperlink network of 118 homeowner forums in Guangzhou to understand collective action behavior among homeowners. Several studies also apply log data (e.g., SMS logs) to explore questions pertaining to language usage patterns of SMS users or internet filtering in China.

Many studies examine digitized data from traditional media outlets. We found 16 articles that look to newspaper data to understand key communication questions such as the role of state propaganda in public discourse, how governments conduct public relations internationally, how news is produced by

journalists and media organizations, and how agenda-setting occurs within a networked media setting. For example, Guo (2019) evaluated 33,875 articles from 291 Chinese official and commercial media outlets during the Two Sessions to assess how coverage differed and how each influenced the other's agenda. Other media types are harder to categorize but no less important. We also saw many studies using computational methods to analyze television programs, commercials, instruction manuals, and Christian songs.

While media data was most common in our dataset, many studies also examine communication phenomena less commonly understood as "media," such as communication within or between institutional organizations, often with the aim of shedding light on questions of corporate social responsibility, organizational partnership networks, and research and development collaboration. These types of studies also analyze political campaigns, survey responses, and opinion polls.

Methods

Computational Chinese communication from our dataset is dominated by two methods: *computer-assisted content analysis* and *social network analysis*. In broad terms, "computer-assisted content analysis" refers to the use of algorithms to study message content, while "social network analysis" refers to the examination of social structures within networks. In our dataset, Mei (2008) was the first study to use computer-assisted content analysis, putting forward an integrated methodology of "commercial and self-developed online content analysis tools" to understand Chinese Internet forums. In succeeding years, methods became more sophisticated. In the 2010s, researchers used a greater proportion of techniques originating in computer science, including natural language processing, or machine learning methods. Nearly all of these studies used text as the main source of data, with a relative dearth of other content types such as images, audios, and videos. This overrepresentation of text data risks overlooking important communication phenomena operating on other modalities and suggests that content analysis of Chinese communication phenomena may benefit from the use of newer computational techniques such as computer vision and signal processing.

Social network analysis, meanwhile, is a relatively broad category which focuses on the social structures embedded in a network and properties of units or groups. In the context of Chinese communication, networks under study vary widely. In our dataset (and especially in the 2010s), we found studies examining netizens, organizations, hyperlinks, newspapers, and many others besides. Additionally, researchers also combine network analysis with content analysis via the introduction of semantic network analysis, which combines the essence of social network analysis and content analysis, to create networks among semantic components to reveal the structures of text.

These two methods—computational-assisted content and network analysis—are often complemented by the use of time series analysis to tease out complicated

questions of causality, influence, and change. For example, Guo (2019) used Granger causality models and unsupervised topic modeling to reveal temporal shifts in news websites' coverage patterns, helping to gain insight on intermedia agenda setting in an online context. In another case, Zeng et al. (2017) combined time series analysis with clustering analysis to understand content moderation strategies on Weibo during the 2015 Tianjin explosion.

Outside of these methods, other techniques highlighted by Hilbert et al. (2019), such as digital simulations, received only limited attention, with only two simulation studies identified in our dataset after 2017. Other computational methods sometimes found in non-Chinese communication research, such as geospatial analysis and computational biology, were not represented at all in our data. These absences, alongside the absence of non-textual modalities, suggests Chinese communication research still has much room for further innovation and exploration.

METHODOLOGICAL CHALLENGES: RETHINKING COMMENSURABILITY, VALIDITY, AND RELIABILITY

Computational methods and online experiments offer unique advantages and disadvantages for understanding communication phenomena. Many of these advantages and disadvantages have been discussed extensively elsewhere (Hilbert et al., 2019; Lazer et al., 2009, 2020; van Atteveldt & Peng, 2018). In this section and the section to follow, we discuss the distinct methodological *challenges* and *opportunities* presented by conducting computational research and online experiments in a Chinese context. We discuss challenges first. While not necessarily unique to China, we believe each of the following challenges is more pronounced in the Chinese context owing to China-specific cultural, political, legal, social, and linguistic factors that introduce additional risk of bias. These challenges highlight the importance of adopting a reflexive approach to the computational and online experimental study of Chinese communication.

Commensurability Between the United States and China

The vast majority of computational communication research and online experiments are conducted in Western contexts. These research sites, and the theories, methods, and conclusions reached within them, are not always commensurable with China. While the challenge of research commensurability across countries is not a new one, computational methods induce additional challenges by adding layers of abstraction between the researcher and the data, rendering these sources of incommensurability less visible. In addition, China contains many significant axes of difference with Western countries, including differences in political systems (authoritarian vs. democratic), language (logographic vs. phonetic), and

culture (collectivistic vs. individualistic). We identify three areas of computational social science and online experiments where concerns about commensurability are most pronounced: *data pre-processing*, *model training*, and *construct validity of online phenomena*.

Data Pre-processing

Computational communication methods often require the alteration of data to render it amenable to algorithmic measurement and analysis. Best practices for these pre-processing, cleaning, and post-processing steps are often developed in Western contexts, inducing design choices which are not optimized for China. In natural language processing, for example, many algorithms create "tokens" based on the assumed atomicity of words separated by spaces (Baden et al., 2022), shorten words based on patterns of lexical base units (called "lemmatization"), and make parsing choices based on an assumption of fixed word order (Goldberg & Elhadad 2013). None of these assumptions hold in Chinese, forcing researchers to either develop new algorithms or risk bias via the misapplication of old ones (Ma et al., 2018). While some of these design risks can be compensated by altering model parameters or via additional pre-processing steps, these steps may introduce new biases—forcing researchers to be even more careful when using methods not originally designed for Chinese data.

Model Training

In recent years, communication research has discovered the power of sophisticated computational models pre-trained on large volumes of data to reduce annotation needs and lower computational costs. Much of this training data is gathered in Western contexts, making it likely that Western-specific assumptions and cultural biases are baked into models, often beyond the awareness of downstream users or even the model's creator herself (Cheng et al., 2023; Liang et al., 2021; Weidinger et al., 2021). Sometimes, these biases are detectable: for example, computer vision algorithms trained on Western faces may underperform on Asian facial features (Khalil et al., 2020; Phillips et al., 2011). Other times, these biases may not be detectable at all, such as the hidden political biases embedded in Large Language Models (LLMs) (Zhou & Zhang, 2024), leading to erroneous results of which we are not aware. In both cases the sheer volume of training data needed for the creation of large-scale models, paired with occasional lack of transparency about which data was used for training, makes it challenging for researchers to detect these sources of incommensurability. Often, Chinese communication researchers are faced with an unwelcome choice: apply pre-trained models with unknown biases to a Chinese context, or use models trained

on less data or with less validation (Huaping et al., 2023). The creation of open-source models trained on Chinese data remains an ongoing challenge in the development of high-performing LLMs, a challenge shared by computational communication researchers.

Construct Validity of Online Phenomena

Prior digital research has developed a vocabulary of correspondences between online phenomena (e.g., "liking" a post) and theoretical constructs (e.g., endorsement of a message). In many cases, these associations have been drawn primarily on the basis of Western platform affordances and user behavior, leading communication research to ascribe theoretical meaning to online behaviors which improperly universalize Western norms. For example, "hyperlinking," which is at base the online affordance of forwarding users between web pages, has been interpreted by computational social scientists as a measure of "influence," "relevance," "credibility," and "self-expression" (Freelon, 2014). In a Chinese online communication environment, which to a greater extent than the West occurs within the walled gardens of digital platforms such as WeChat or Weibo which discourage and sometimes outright prohibit the use of hyperlinks that direct users off-platform (Jia & Winseck, 2018), hyperlinks can and perhaps should be associated with a more rigorous set of theoretical constructs than in the West. In other cases, the online affordance itself may have no obvious communication equivalent in the Western world, such as "daka" culture (the phenomenon of digitally marking that one has visited a place) (He, 2024). In any of these cases, China's distinct affordances and platform restrictions ought to precipitate a reconsideration of theoretical constructs.

At other times, China's unique cultural context may threaten construct validity from the reverse direction: online behavior may differ so drastically as to give the same affordance a different meaning. Consider online fandom. In Western contexts, fandom has typically been associated with online participatory cultures (Jenkins, 2012) situated oppositional to mainstream political culture (e.g., K-pop fans opposing Donald J. Trump's reelection). In China, strict state control of both celebrities and fans may lead fans to engage in similar-seeming online activities (e.g., "liking" certain posts) but with radically different intentions and consequences. For example, combining computational analysis of celebrity fans' social media activities and qualitative interviews with fans, Lu (2023) uncovered a "performative" dimension to online engagement with celebrity-endorsed propaganda, revealing a completely different meaning to the "commenting" and "resharing" of content (Lu, 2023). Computational communication researchers must devise new theoretical constructs with sensitivity to both distinct platform affordances and distinct platform behavior. Importing communication constructs from Western contexts, if improperly applied, may lead to a serious mischaracterization of the phenomena under study.

Validity and Reliability

Incommensurability of measures, data, and constructs are challenges that arise when researchers import computational communication research practices, developed largely outside of China, directly into a Chinese context. However, other challenges arise less from importation of foreign practices than from indigenous characteristics of the Chinese system itself. To be clear, commensurability challenges (outside-in) and challenges from the Chinese context itself (inside-in) can also be understood from the other perspective – imported NLP algorithms' difficulty with Chinese logographic scripts (commensurability), for example, can also be understood as an indigenous feature of the Chinese system itself. We divide the two because their challenges to communication research are subtly different. Challenges of commensurability are often detectable: a lower F1 score, for instance, or an error message when parsing Chinese text. Indigenous challenges, on the other hand, are more subtle, often requiring a top-to-bottom reconsideration of research design, and greater vigilance by the researcher. Moreover, these indigenous challenges more directly threaten research *validity* (whether computational techniques measure the concepts authors intend to measure) and *reliability* (whether computational techniques can consistently measure the concepts of interest) (Chen et al., 2024). Below we outline the three biggest indigenous threats to computational communication research in China: *state regulation*, *algorithmic influence*, and *shifting online behaviors*.

State Regulation

The Chinese government operates the largest and most sophisticated apparatus for the control and influence of online communication the world has ever seen (King et al., 2013). For computational communication researchers, these myriad and often hidden forms of state regulation complicate both the gathering and subsequent interpretation of online data and online experiments. When examining the production of speech, for example, the ubiquity of digital repression (Earl et al., 2022) and the possibility of offline reprisal (DeButts & Pan, 2024) may lead users to engage in self-censorship or ironic speech less commonly seen in democratic societies (Chung & Fu, 2022; Nordin & Richaud, 2014). At other times, when computational researchers seek to gather speech that has already been produced, government propaganda or platform censorship (King et al., 2013; Lu & Pan, 2021) may result in digital trace data containing pro-government biases (King et al., 2017) or an artificial absence of threatening counter-hegemonic discourse (Zeng et al., 2017). While some steps can be taken (Shen & Truex, 2021) to account for the biases introduced by government regulation (and using computational methods to study government intentions *by way of* examining these regulations is a fruitful opportunity for computational research in China, as discussed in the "Opportunities" section), the true effect of state regulations on data used in computational communication research is hard to assess.

State regulation also affects data availability. The Chinese government is reluctant to allow unfettered research access to Chinese citizen data, and also reluctant to allow the outbound transfer of Chinese citizen data to other countries (Perez, 2022). This protectionism poses a direct threat to the validity of communication research performed in China. Online experiments, for example, usually seek to obtain random samples from a population. The Chinese government rarely allows for random population sampling, exacerbating the threat of selection bias (i.e. that researchers select users for study on the basis of an unobserved shared characteristic which is correlated with the treatment condition). Computational research, which requires large-scale datasets of user speech or behavior, runs the risk of violating national security laws which forbid the exportation of user data outside the country. Online field experiments in China, which are often performed in partnership with Chinese universities so as to gain access to research subjects, often require approval from government or school officials reluctant to permit "sensitive" research inquiries or treatment conditions. Moreover, all of these factors are constantly changing: what was permissible two years ago may not be permissible today, threatening the reliability and sustainability of computational and experimental research performed online in China. These political sensitivities require researcher attention at all stages of the research process, from data collection to interpretation to analysis.

Algorithmic Influence

A second challenge for computational communication research in China is unobservable variance in platform algorithms. One source of variance is hidden state directives. China's social media market, unlike most other countries in the world, is dominated by domestic champions (Pan, 2017). These domestic platforms are thus highly responsive to formal and informal government instructions, many of which are not visible to outside observers (Creemers, 2017). Weibo's trending topics, for example, exerts a significant impact on online user behavior and are also tweaked at the behest of the government (Yang & Peng, 2022). Online censorship, too, is carried out primarily at the corporate level, resulting in platform-specific keyword bans based on platform-level differences in risk assessment, risk tolerance and censorship capability (Knockel et al., 2017). Another source of variance is platforms' divergent commercial incentives, such as whether they earn money from subscriptions or user data. These divergences also affect the types of affordances they offer, behavior they encourage, and data they record, each of which affects the types of questions computational communication researchers can profitably ask and answer (Chen et al., 2024). A third source of variance is algorithmic recommendation systems, which are usually distinct to each platform and may be tailored also on the basis of each user. These sources of variance are compounded by the relative intransparency of Chinese digital platforms, which sometimes make sudden and unannounced

changes to data gathering processes, threatening the reliability of computational research (Howison et al., 2011).

Shifting Online Behaviors

As elsewhere in the digital world, online behaviors in China are not necessarily stable over time. As new users enter and exit online populations, speech and other behavioral patterns shift, complicating the validity and reliability of computational results (Salganik, 2019). New users are sometimes associated with online slang and other neologisms (Danescu-Niculescu-Mizil et al., 2013), which may weaken the validity of dictionary-based natural language processing techniques and require the creation of new algorithms for detection (Pei et al., 2019). Even if populations remain stable, behaviors may change – for example, changing political conditions, such as greater scrutiny of online influencers, may lead users to self-censor (Shen & Truex, 2021), and cultural proximity influences longitudinal browsing habits (Taneja & Wu, 2014). All of these may potentially lead researchers to errant conclusions when datasets are less up-to-date. Many of these concerns apply to online experiments as well. For example, as more people use Rednote (Xiaohongshu) as their first-choice search platform (rather than Baidu or Weibo) in recent years, new experiments occurring on outdated platforms may draw from unrepresentative populations or unrepresentative user behavior, threatening the external validity of results. These online shifts are often subtle and indirectly observable, making it challenging for researchers to properly account for them.

Challenges to computational communication research and online field experiments are endemic. Some challenges arise from the importation of disciplinary conventions into a Chinese context, such as data pre-processing, model training, and construct validity. These challenges may also threaten research validity and reliability, but they are often observable. Other challenges arise more directly from features of the Chinese research context itself, including state regulation, divergent platform incentives, and changing online behaviors. These challenges are multifaceted and often less visible. In all cases, computational research and field experiments in China would benefit from explicitly addressing these challenges, with the goal of eventually developing China-specific research practices to account for them.

OPPORTUNITIES: DATA ABUNDANCE, DYNAMIC ONLINE ACTIVITIES, AND THEORETICAL OPPORTUNITIES

China also offers a tremendous opportunity for researchers conversant in computational communication techniques and online experiment best practices. The country's past thirty years of internet connectivity has been characterized by

rapid digital growth, for example in mobile payments (Fan et al., 2018) and within the domestic app ecosystem (Muise et al., 2022), as a growing proportion of citizen behavior is logged and tracked online. This relative bounty of online data, alongside a dynamic online environment, makes possible a truly kaleidoscopic diversity of communication research (Trilling et al., 2024). The relative dearth of China-specific media theories renders these inquiries all the more pressing and valuable.

Ubiquity of Online Data

In the past thirty years, an ever-growing proportion of Chinese life takes place online. A mixture of political, commercial, and social incentives have conspired to make this happen. Politically, the Chinese central government has paired market protectionism (Yalcintas & Alizadeh, 2020), which has facilitated the rise of domestic internet champions, with campaigns to digitize government services, in part to resolve principal-agent problems (Qiaoan & Teets, 2020). The result is the widespread provision of e-government services at all administrative levels (Zhou, 2004), as well as a set of incentives for lower-level governments that strongly encourages the creation of more digital data (Gao, 2020). Commercially, China has benefited from "leapfrog" growth in internet services, coinciding with an emergence of the world's largest population of digital payments users (Han & Wang, 2021). At the same time, intense competition between companies (including state-backed companies) has created some of the world's most innovative technology companies, including Alibaba and Tencent (Jia & Winseck, 2018), which together have cultivated a flourishing domestic ecosystem of mobile apps and user retention on these platforms. Finally, from a social perspective, Chinese citizens have consistently seen the internet as a place for the pursuit of civic action, political accountability, economic opportunity, and community and belonging (Chen & Li, 2024; Yang, 2009). Recent research has also highlighted the limitations of internet censorship in fully isolating Chinese voices from the rest of the world, with substantial transnational information flows continue within the Chinese internet (Lu et al., 2024).

The result of these combined forces is a massive effluence of data generated constantly by a wide cross-section of Chinese society, from youth to elderly, rich to poor. The space may soon become more crowded – the advent of generative artificial intelligence (AI) may produce even more types of data, including algorithmically-authored text and videos, in turn requiring new methods and theories to analyze. The sheer volume of all this data makes possible a wide array of computational communication questions, such as assessing digital repression, human–bots interactions, and online advertisements. The size of China's internet population and the quantity of their online activities are a significant opportunity for computational communication researchers.

Diversity of Online Activities in China

In addition to the quantity of data, China also offers an impressive *diversity* of activities taking place online. According to CNNIC 52nd Statistical Report on China's Internet Development (2023), China's internet currently comprises over 3.83 million websites and 2.6 million active apps. Moreover, the functionality of these apps is rapidly developing: early CNNIC reports primarily feature apps with basic functions (e.g., instant messaging and search engine), while later reports highlight a wide range of functionalities, including short-video creation, online gaming, e-reading, live-streaming, ride-sharing, and online medical services. Perhaps unsurprisingly, user numbers have increased apace, with volume increasing steadily over the last few decades. Muise et al. (2022) finds that a subsample of Chinese mobile phone users spent on average 40.6 hours each week using their phones, interacting with a wide variety of applications including mobile gaming, music, and social media. Such variety is perhaps in part by the growth in digital super-platforms like WeChat, which can administer a growing slice of user needs—from messaging to purchasing movie tickets to web-browsing—without users ever leaving the platform. That these activities all occur within the same mobile app ecosystem, and often at the same time, makes possible new lines of theoretical inquiry across unique combinations of subfields: interpersonal communication and marketing, for example, or political communication and public health.

Building China-specific Media Theories

As stated in the previous section, the Western derivation of many communication methods, models, and theories is a significant challenge for computational communication in China. However, that relative dearth of China-specific communication theories is also an opportunity. Research examining China's communication environment has uncovered phenomena with scarce comparability in Western contexts, such as how government officials enlist clickbait in the pursuit of citizen attention (Lu & Pan, 2021), how cultural industries operate under close supervision (Fang, 2024), and how government and media compete and collaborate to apply scrutiny to government misbehavior (Repnikova, 2017). There is much scope for further exploration of these phenomena, and thus developing more theories to explain these phenomena as well as the Chinese Internet (Yang, 2012). For example, scholars have only recently begun to examine whether negatively framed information, which in Western countries has been shown to be perceived as more credible than equivalent positively framed information, also holds in a Chinese context (Yao & Li, 2024), or the applicability of agenda-setting in a Chinese media ecosystem (Guo & Zhang, 2020).

Much of modern Chinese life occurs online. The *volume* and *diversity* of these activities, coupled with the relative absence of *China-specific communication theories* to characterize them, excavates a wide space for theoretical innovation

using computational methods. Put succinctly, computational communication research on China can capture a rich slice of the human experience simply because more of daily life occurs there. For communication researchers who are interested in varieties of the human experience, this combination presents welcome opportunities.

DISCUSSION

Computational methods and online experiments, much heralded for their methodological strength and versatility, can offer significant insights into Chinese digital media. This chapter initiates a conversation about the current state of those methods when applied to China, as well as challenges and opportunities pertaining to their use. Our empirical analysis of the past 20 years of communication research reveals an increasing deployment of computational methods; online experiments, however, remain sparsely adopted. We call, therefore, for a broader adoption of online experiments in exploring Chinese communication phenomena, as their scalability and internal validity bring unique advantages.

Our empirical findings also shed light on the dominant subfields, data types, and methodologies prevalent in computational studies of Chinese digital media. These findings suggest ample scope for the development of both alternative digital data types and new computational methods. For example, recent restrictions on API usage of Western media platforms like Twitter and Reddit underscore the danger of over-reliance on API-driven data collection (Freelon, 2018). Echoing recent reflections by Ohme et al. (2023), we encourage more use of user-centric data rather than platform-centric data to capture the nuanced behaviors and dynamics of digital media activities in China. For instance, Reeves et al. (2021) demonstrates how screen tracking data can capture granular mobile phone activities, and Muise et al. (2022) have taken early steps in applying such methods to China users' mobile phone data. In addition, simulation-based methods offer a viable and underused approach that can compensate for some of the accessibility, validity, and reliability challenges of large-scale digital trace data, and LLMs may further mitigate these challenges by generating synthetic data and supporting simulations.

We hope that our reflections on the challenges to commensurability and validity/reliability of computational methods and digital experiments in China can provide future researchers with a first few considerations when contemplating the use of these methods. Commensurability challenges, particularly those related to conceptual constructs and computational processes (algorithmic pre-processing and model training), represent a warning against the uncritical application of tools or concepts developed in non-Chinese contexts to the study of Chinese digital media. At the same time, factors like state regulation, divergent platform incentives, and changing online behaviors present significant challenges to

computational methods' validity and reliability. These challenges, often more subtle and less observable than commensurability issues, demand careful consideration and strategic planning. We suggest a reflexive approach in the design and implementation of future computational research and online experiments to mitigate the impact of these challenges. In addition, we also encourage researchers to explicitly articulate these challenges when discussing their research to help readers understand the implications of their results.

Despite these multifaceted challenges, we also recognize distinctive opportunities for computational and online experimental research in the study of Chinese digital media. The dynamism of Chinese digital life generates abundant and multifaceted digital traces, offering a panoramic view of user behaviors and media interactions. At the same time, computational and experimental researchers can treat the relative dearth of China-specific media theories as an opportunity to develop theory-driven research that can contribute to the understanding of Chinese digital media both theoretically and empirically.

Based on these reflections, we conclude this chapter by providing two general recommendations for future researchers interested in employing computational and online experimental methods to investigate Chinese digital media. First, we encourage the adoption of data and method triangulation (Mathison, 1988) as a hedge against the risks to validity and reliability of singular data sources or methods. For example, as Peng et al. (2023) suggested, researchers can combine computational analysis of objective visual features (e.g., image brightness) with online crowdsourced ratings of subjective persuasive strategies (e.g., humor and emotions) to provide a comprehensive understanding of visual misinformation feature patterns. Researchers can also design factorial experiments to manipulate prevalent features, assessing the impact of each on credibility perception or persuasive effects. These triangulation strategies are not only applicable to misinformation studies, but a wide variety of media effects studies besides in the context of Chinese communication research.

In addition to these multimethod studies, researchers can also combine computational and online experimental methods with qualitative methods to conduct mixed methods research. Such an approach doesn't just allow for different data types – it entails different worldviews and epistemologies (Venkatesh et al., 2013). There are a few promising studies employing this mixed approach: Zeng et al. (2022) used both qualitative and computational methods to analyze official and public discourse about artificial intelligence, while Lu and Pan (2021) combined computational analysis of WeChat government article titles with ethnographic fieldwork with Chinese local propaganda teams to understand how local governments compete for online attention. We expect studies that leverage different research methods and strategies will contribute even more to our comprehensive knowledge of Chinese digital media.

Finally, we also issue a call for greater collaboration. The combined challenge of data inaccessibility and data volume necessitate collaborations across institutions and regions. In many cases, this collaboration may revolve around data

collection efforts. Chinese social media platforms like Weibo, WeChat, Douyin (Chinese TikTok), and Rednote are of great interest to researchers but often quite challenging to study, as data is not easy to gather. Communication research as a whole would benefit from sharing knowledge on how to access these platforms and collect data from within them. These data-gathering and data-sharing infrastructures can be paired, when possible, with open science principles. Such an approach ensures that the unique capabilities of computational methods are shared as widely as possible, and that the greatest number of researchers—both inside China and outside it—can enjoy their benefits.

ACKNOWLEDGMENT

We gratefully acknowledge Dr. Kaiping Chen for generously sharing the data used in this chapter.

REFERENCES

Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures*, *16*(1), 1–18.

Bresnahan, M. J., & Zhuang, J. (2016). Effects of normative antismoking messages in China—a brief report. *Health Communication*, *31*(3), 374–378.

Chen, S., & Li, Y. (2024). Accountability from cyberspace? Scandal exposure on the internet and official governance in China. *Political Science Research and Methods*, *12*(1), 130–145.

Chen, K., Lu, Y., & Wang, Y. (2024). Unraveling China's digital traces: Evaluating communication scholarship through a sociotechnical lens. *Chinese Journal of Communication*, 1–24.

Chen, Y., Lu, J., & Wu, X. (2023). China' as a 'black box?' Rethinking methods through a sociotechnical perspective. *Information, Communication & Society*, 1–17.

Chen, L., & Yang, X. (2018). Using EPPM to evaluate the effectiveness of fear appeal messages across different media outlets to increase the intention of breast self-examination among Chinese women. *Health Communication*, *34*(11), 1369–1376.

Cheng, M., Durmus, E., & Jurafsky, D. (2023). *Marked personas: Using natural language prompts to measure stereotypes in language models*. arXiv preprint arXiv:2305.18189.

Chung, R. W. M., & Fu, K. W. (2022). Tweets and memories: Chinese censors come after me. Forbidden voices of the 1989 Tiananmen Square massacre on Sina Weibo, 2012-2018. *Journal of Contemporary China*, *31*(134), 319–334.

CNNIC. (2023). *Statistical report on internet development in China*. China Internet Network Information Center.

Creemers, R. (2017). Cyber China: Upgrading propaganda, public opinion work and social management for the twenty-first century. *Journal of Contemporary China*, *26*(103), 85–100.

Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., & Potts, C. (2013, May). No country for old members: User lifecycle and linguistic change in online communities. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 307–318).

DeButts, M., & Pan, J. (2024). Reporting after removal: The effects of journalist expulsion on foreign news coverage. *Journal of Communication*, *74*, 273–286.
Earl, J., Maher, T. V., & Pan, J. (2022). The digital repression of social movements, protest, and activism: A synthetic review. *Science Advances*, *8*(10), eabl8198.
Fan, J., Shao, M., Li, Y., & Huang, X. (2018). Understanding users' attitude toward mobile payment use: A comparative study between China and the USA. *Industrial Management & Data Systems*, *118*(3), 524–540.
Fang, J. (2024). The culture of censorship: State intervention and complicit creativity in global film production. *American Sociological Review*, *89*(3), 00031224241236750.
Feng, G. C. (2019). A comparative study of the online film ratings of US and Chinese audiences: An analytical approach based on big data. *International Communication Gazette*, *81*(3), 283–302.
Freelon, D. (2014). On the interpretation of digital trace data in communication and social computing research. *Journal of Broadcasting & Electronic Media*, *58*(1), 59–75.
Freelon, D. (2018). Computational research in the post-API age. *Political Communication*, *35*(4), 665–668.
Gao, X. (2020). State-society relations in China's state-led digitalization. *China Review*, *20*(3), 1–12.
Goldberg, Y., & Elhadad, M. (2013). Word segmentation, unknown-word resolution, and morphological agreement in a Hebrew parsing system. *Computational Linguistics*, *39*(1), 121–160.
Guo, L. (2019). Media agenda diversity and intermedia agenda setting in a controlled media environment: A computational analysis of China's online news. *Journalism Studies*, *20*(16), 2460–2477.
Guo, L., & Zhang, Y. (2020). Information flow within and across online media platforms: An agenda-setting analysis of rumor diffusion on news websites, Weibo, and WeChat in China. *Journalism Studies*, *21*(15), 2176–2195.
Han, P., & Wang, Z. (2021). *Technology adoption and leapfrogging: Racing for mobile payments*. https://ssrn.com/abstract=3818168
He, Y. (2024). Exploring linguistic innovations within Chinese internet slang in language and social media. *Journal of Linguistics and Communication Studies*, *3*(1), 29–34.
Hilbert, M., Barnett, G., Blumenstock, J., Contractor, N., Diesner, J., Frey, S., González-Bailón, S., Lamberson, P. J., Pan, J., Peng, T.-Q., Shen, C., Smaldino, P. E., van Atteveldt, W., Waldherr, A., Zhang, J., & Zhu, J. J. (2019). Computational communication science: A methodological catalyzer for a maturing discipline. *International Journal of Communication*, *13*, 3912–3934.
Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, *12*(12), 2.
Huang, R., & Sun, X. (2016). Dynamic preference revelation and expression of personal frames: How Weibo is used in an anti-nuclear protest in China. *Chinese Journal of Communication*, *9*(4), 385–402.
Huang, Z. A., & Wang, R. (2019). Building a network to "tell China stories well": Chinese diplomatic communication strategies on Twitter. *International Journal of Communication*, *13*, 2984–3007.
Huaping, Z., Linhan, L., & Chunjin, L. (2023). ChatGPT performance evaluation on Chinese language and risk measures. *Data Analysis and Knowledge Discovery*, *7*(3), 16–25.
Jenkins, H. (2012). *Textual poachers: Television fans and participatory culture*. Routledge.

Jia, L., & Winseck, D. (2018). The political economy of Chinese internet companies: Financialization, concentration, and capitalization. *International Communication Gazette*, *80*(1), 30–59.

Khalil, A., Ahmed, S. G., Khattak, A. M., & Al-Qirim, N. (2020). Investigating bias in facial analysis systems: A systematic review. *IEEE Access*, *8*, 130751–130761.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, *107*(2), 326–343.

King, G., Pan, J., & Roberts, M. E. (2017). How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, *111*(3), 484–501.

Knockel, J., Ruan, L., & Crete-Nishihata, M. (2017). Measuring decentralization of Chinese keyword censorship via mobile games. In *7th USENIX Workshop on Free and Open Communications on the Internet (FOCI 17)*.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational social science. *Science*, *323*(5915), 721–723.

Lazer, D. M., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., González-Bailón, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science*, *369*(6507), 1060–1062.

Li, Y., Rau, P. L. P., Li, H., & Maedche, A. (2017). Effects of a dyad's cultural intelligence on global virtual collaboration. *IEEE Transactions on Professional Communications*, *60*(1), 56–75.

Liang, P. P., Wu, C., Morency, L. P., & Salakhutdinov, R. (2021, July). Towards understanding and mitigating social biases in language models. In *International conference on machine learning* (pp. 6565–6576). PMLR.

Lu, Y. (2023). *Performative propaganda engagement: How celebrity fandom engages with state propaganda in China* (Doctoral dissertation). Stanford University.

Lu, J., & Li, B. (2015). Media representation of mobile technologies: A content analysis of television commercials. *International Journal of Mobile Communications*, *13*(3), 229–243.

Lu, Y., & Pan, J. (2021). Capturing clicks: How the Chinese government uses clickbait to compete for visibility. *Political Communication*, *38*(1–2), 23–54.

Lu, Y., Schaefer, J., Park, K., Joo, J., & Pan, J. (2024). How information flows from the world to China. *The International Journal of Press/Politics*, *29*(2), 305–327.

Ma, J., Ganchev, K., & Weiss, D. (2018). *State-of-the-art Chinese word segmentation with Bi-LSTMs*. arXiv preprint arXiv:1808.06511.

Mathison, S. (1988). Why triangulate? *Educational Researcher*, *17*(2), 13–17.

Mei, W. (2008). Measuring political debate on the Chinese internet forum. *Javnost-The Public*, *15*(2), 93–110.

Muise, D., Lu, Y., Pan, J., & Reeves, B. (2022). Selectively localized: Temporal and visual structure of smartphone screen activity across media environments. *Mobile Media & Communication*, *10*(3), 487–509.

Muise, D., & Pan, J. (2019). Online field experiments. *Asian Journal of Communication*, *29*(3), 217–234.

Nordin, A., & Richaud, L. (2014). Subverting official language and discourse in China? Type river crab for harmony. *China Information*, *28*(1), 47–67.

Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., & Robinson, T. N. (2023). Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking. *Communication Methods and Measures*, *18*(2), 1–18.

Pan, J. (2017). How market dynamics of domestic and foreign social media firms shape strategies of internet censorship. *Problems of Post-Communism*, *64*(3–4), 167–188.

Pei, Z., Sun, Z., & Xu, Y. (2019, November). Slang detection and identification. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)* (pp. 881–889).

Peng, Y., Lu, Y., & Shen, C. (2023). An agenda for studying credibility perceptions of visual misinformation. *Political Communication*, *40*(2), 225–237.

Perez, C. (2022). *Why China's new data security law is a warning for the future of data governance*. Foreign Policy. https://foreignpolicy.com/2022/01/28/china-data-governance-security-law-privacy/

Phillips, P. J., Jiang, F., Narvekar, A., Ayyad, J., & O'Toole, A. J. (2011). An other-race effect for face recognition algorithms. *ACM Transactions on Applied Perception (TAP)*, *8*(2), 1–11.

Qiaoan, R., & Teets, J. C. (2020). Responsive authoritarianism in China–a review of responsiveness in Xi and Hu administrations. *Journal of Chinese Political Science*, *25*(1), 139–153.

Reeves, B., Ram, N., Robinson, T. N., Cummings, J. J., Giles, C. L., Pan, J., Chiatti, A., Cho, M. J., Roehrick, K., Yang, X., Gagneja, A., Brinberg, M., Muise, D., Lu, Y., Luo, M., Fitzgerald, A., & Yeykelis, L. (2021). Screenomics: A framework to capture and analyze personal life experiences and the ways that technology shapes them. *Human–Computer Interaction*, *36*(2), 150–201.

Repnikova, M. (2017). *Media politics in China: Improvising power under authoritarianism*. Cambridge University Press.

Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.

Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big data, digital media, and computational social science: Possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 6–13.

Shen, X., & Truex, R. (2021). In search of self-censorship. *British Journal of Political Science*, *51*(4), 1672–1684.

Taneja, H., & Wu, A. X. (2014). Does the great firewall really isolate the Chinese? Integrating access blockage with cultural factors to explain web user behavior. *The Information Society*, *30*(5), 297–309.

Trilling, D., Araujo, T., Kroon, A., Möller, A. M., Strycharz, J., & Vermeer, S. (2024). *Computational communication science in a digital society*. In T. Araujo & P. Neijens (Eds.), *Communication Research into the Digital Society: Fundamental insights from the Amsterdam School of Communication Research*. (pp. 247–263). Amsterdam University Press.

van Atteveldt, W., Margolin, D., Shen, C., Trilling, D., & Weber, R. (2019). A roadmap for computational communication research. *Computational Communication Research*, *1*(1), 1–11.

van Atteveldt, W., & Peng, T. Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, *12*(2–3), 81–92.

Veneti, A., Karadimitriou, A., & Poulakidakos, S. (2016). Media ecology and the politics of dissent: Representations of the Hong Kong protests in the Guardian and China Daily. *Social Media+ Society*, *2*(3).

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, *37*(1), 21–54.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., . . ., & Gabriel, I. (2021). *Ethical and social risks of harm from language models*. arXiv preprint arXiv: 2112.04359.

Wu, D., & Yang, A. (2017). China's public diplomatic networks on the Ebola issue in West Africa: Issues management in a network society. *Public Relations Review*, *43*(2), 345–357.

Yalcintas, A., & Alizadeh, N. (2020). Digital protectionism and national planning in the age of the internet: The case of Iran. *Journal of Institutional Economics*, *16*(4), 519–536.

Yang, G. (2009). *The power of the Internet in China: Citizen activism online*. Columbia University Press.

Yang, G. (2012). A Chinese internet? History, practice, and globalization. *Chinese Journal of Communication*, *5*(1), 49–54.

Yang, T., & Peng, Y. (2022). The importance of trending topics in the gatekeeping of social media news engagement: A natural experiment on Weibo. *Communication Research*, *49*(7), 994–1015.

Yao, S., & Li, H. (2024). Equivalence framing and its effects on truth judgments: Evidence from China. *Chinese Journal of Communication*, 1–17.

Zeng, J., Chan, C. H., & Fu, K. W. (2017). How social media construct "truth" around crisis events: Weibo's rumor management strategies after the 2015 Tianjin blasts. *Policy & Internet*, *9*(3), 297–320.

Zeng, J., Chan, C. H., & Schäfer, M. S. (2022). Contested Chinese dreams of AI? Public discourse about artificial intelligence on WeChat and People's Daily Online. *Information, Communication & Society*, *25*(3), 319–340.

Zhao, T. (2023). Knowledge mapping and empirical framework of computational communication research: A comparative analysis of 12,991 core Chinese and international publications [计算传播研究的知识地图与实证框架——基于12991篇中外核心文献的对比分析]. *Journal of Journalism and Communication Research [新闻与传播研究]*, *30*(3), 40–59, 126–127.

Zhong, Z., Zhou, J., & Su, S. (2023). A comparison of computational communication research in Chinese and English: Topics, methods, and theories [中英文计算传播研究的比较: 议题、方法与理论]. *Journal of Global Media Studies [全球传媒学刊]*, *10*(1), 56–76.

Zhou, X. (2004). E-Government in China: A content analysis of national and provincial web sites. *Journal of Computer-Mediated Communication*, *9*(4), JCMC948.

Zhou, D., & Zhang, Y. (2024). Political biases and inconsistencies in bilingual GPT models—the cases of the US and China. *Scientific Reports*, *14*(1), 25048.

APPENDIX A KEYWORD LIST

We developed a keyword list to extract targeted article abstracts through querying two sets of keywords: (1) those indicating computational studies in Chinese communication, such as "machine learning" AND "China"; and (2) those indicating experimental studies in Chinese communication, like "random assignment" AND"China."

Computational studies	Experimental studies	China-related keywords
Computational	Experiment*	China
API machine learning	Randomiz*	Chinese
Computational	RCT quasi*	Beijing
Data mining	Causal	Sino
Data-mining	Stimuli	CCP
Deep learning	Factorial	Mandarin
Network analysis	Pilot stud*	PRC
Big data	Virtual reality	Hong Kong
Natural language processing	VR eye-track*	Taiwan
Hyperlink	Eye track*	Weibo
NLP agent-based	A/B game theory	Baidu
Python	Trial	WeChat
" R " geospatial	Treatment	Douyin
simulation	Control group	Douban
Bayesian	Pretest	Zhihu
Frequentist	Post test	
Artificial intelligence	Longitudinal study	
Supervised learning	Intervention	
Unsupervised learning	Repeated measures	
Graph theory		
Computer vision		
Text analy*		
Time series*		
Digital tool*		
Semantic tagging analys*		
Quantitative content analys*		