

## Original Paper

# Spatial Heterogeneity of Digital HIV/AIDS Stigma in China: Cross-Sectional Ecological Study

Pu Zhang<sup>1</sup>, MPhil; Yiliang Li<sup>2</sup>, MSc; Zheng Wei<sup>3</sup>, PhD; Na Jiang<sup>1,4</sup>, PhD; Pan Hui<sup>1,3,5</sup>, PhD

<sup>1</sup>The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, Guangdong, China

<sup>2</sup>Wuhan University, Wuhan, Hubei, China

<sup>3</sup>The Hong Kong University of Science and Technology, Hong Kong, China (Hong Kong)

<sup>4</sup>Guangdong Provincial Key Laboratory of Smart and Safety City Planning and Monitoring, Guangzhou, Guangdong, China

<sup>5</sup>University of Helsinki, Uusimaa, Helsinki, Finland

## Corresponding Author:

Na Jiang, PhD  
The Hong Kong University of Science and Technology (Guangzhou)  
1st, Duxue Road  
Guangzhou, Guangdong 511453  
China  
Phone: 86 19575360103  
Email: [najiang@hkust-gz.edu.cn](mailto:najiang@hkust-gz.edu.cn)

## Abstract

**Background:** HIV/AIDS stigma remains a major barrier to public health intervention and social inclusion. In China, short-video platforms, such as Douyin, have become important spaces where people living with HIV/AIDS share daily-life narratives and receive public responses. However, little is known about how public responses to these narratives vary geographically, or how regional sociodemographic and epidemiological contexts are associated with digital stigma and support.

**Objective:** This study aimed to examine the thematic structure, spatial heterogeneity, and macro-level correlates of public responses to daily-life narratives shared by people living with HIV/AIDS on Douyin in China, with a focus on spatially varying correlational patterns of digital HIV/AIDS stigma.

**Methods:** We constructed a corpus of 174,394 public comments from Douyin videos created by people living with HIV/AIDS. DeepSeek R1 671B was used to classify comments into 9 substantive themes and 1 residual “others” category. Provincial topic indices were calculated by normalizing thematic comment counts by mobile internet access users. We then combined the least absolute shrinkage and selection operator feature selection, ordinary least squares, and spatial econometric models, residual spatial autocorrelation diagnostics, false discovery rate correction, and Shapley Additive Explanations (SHAP)–geographic information system (GIS) interpretation to examine global associations and localized variable contributions across 31 provincial-level units in mainland China.

**Results:** Moral judgment and attribution of blame were the most prevalent substantive theme (41,663/174,394, 23.89%), followed by disease knowledge and transmission discussion (23,344/174,394, 13.39%), and support, encouragement, and humanitarian care (18,101/174,394, 10.38%). Spatial analysis showed pronounced geographic heterogeneity in both stigma- and support-oriented public responses. HIV incidence was positively associated with disease knowledge, social impact discussion, moral judgment, and support, while gender ratio was significantly associated with moral judgment and related evaluative themes. SHAP-GIS further revealed stronger positive contributions of HIV incidence to moralizing discourse in several southwestern provinces and a marked north-south pattern in the contribution of gender ratio.

**Conclusions:** Digital public responses to daily-life narratives shared by people living with HIV/AIDS in China are geographically uneven and associated with place-based sociodemographic and epidemiological contexts.

*J Med Internet Res* 2026;28:e93007; doi: [10.2196/93007](https://doi.org/10.2196/93007)

**Keywords:** HIV/AIDS; stigma; infoveillance; short-video platforms; large language models; spatial analysis; digital health

## Introduction

### Background

AIDS, caused by HIV, remains a formidable challenge at the intersection of public health and societal behavior. By 2024, approximately 40.8 million people were living with HIV globally, with 630,000 AIDS-related deaths annually, underscoring a persistent crisis that biomedical advances alone cannot resolve [1]. In mainland China, the number of people living with HIV/AIDS exceeds 1 million, with rising diagnosis and mortality rates [2]. These epidemiological trends are inextricably linked to sociobehavioral determinants (eg, social stigma, health literacy deficits, and psychological barriers [3]), which not only undermine the well-being of people living with HIV/AIDS but also obstruct systemic efforts to achieve the “95-95-95” public health targets [4,5].

The rapid evolution of computer-mediated communication has fundamentally reshaped how the public consumes and responds to health-related narratives. Short-form video platforms, exemplified by Douyin (the Chinese version of TikTok), have transitioned from entertainment hubs into primary sociotechnical landscapes for information exchange among hundreds of millions of users [6,7]. The short-video format, characterized by user-generated videos typically under 60 seconds with algorithm-driven content distribution, constitutes the primary medium through which these interactions occur on Douyin [8]. For people living with HIV/AIDS, these platforms offer a unique digital affordance for self-disclosure, peer support, and the building of psychological resilience [9]. However, the digital environment is a double-edged sword; the anonymity and high-speed virality of Douyin can facilitate online disinhibition, leading to the rapid proliferation of stigmatizing content and the reinforcement of negative stereotypes [10,11].

Despite the proliferation of HIV/AIDS-related user-generated content, a significant research gap remains regarding the digital determinants of health [12,13], particularly in relation to how online expressions of HIV/AIDS stigma and support are shaped by place-based contexts. In this study, digital HIV/AIDS stigma refers to online expressions of social disapproval, blame attribution, and negative stereotyping toward people living with HIV/AIDS, whereas digital support refers to expressions of empathy, encouragement, and humanitarian care. Specifically, it remains unclear how macro-environmental correlates modulate online behavior [14]. While individual-level psychology is well documented, we lack systematic evidence on how broader regional socioeconomic and epidemiological contexts influence the way netizens interpret and react to narratives of people living with HIV/AIDS [15-17]. Drawing on the concept of infoveillance, we regard comments on Douyin videos created by people living with HIV/AIDS as observable digital traces of public responses to HIV/AIDS narratives [18]. These responses range from moral condemnation, fear-based distancing, and blame attribution to informational exchange and humanitarian support. Understanding this spatial nonstationarity in human behavior, where exposure to similar digital content

may elicit divergent responses depending on the user’s geographic and social substrate, is essential for advancing theories of risk perception and digital health governance [5, 19]. Detailed examples and coding criteria, including the distinction between stigmatizing comments and negative in tone but factually accurate statements, are provided in [Multimedia Appendix 1](#).

To address these gaps, this study uses a large-scale dataset of 174,394 public comments collected from Douyin. We integrate large language models (LLMs) for automated thematic classification with spatial econometric modeling and machine learning interpretability (Shapley Additive Explanations [SHAP]) to decode the complex relationship between regional context and online health attitudes. Specifically, this study is guided by the following 4 research questions (RQs):

- RQ1: What are the primary thematic dimensions and behavioral patterns in the public’s digital discourse regarding the daily-life narratives of people living with HIV/AIDS on short-video platforms?
- RQ2: How is the intensity of these digital themes distributed geographically across China’s provinces, and are there significant spatial clusters of stigma or support?
- RQ3: To what extent are provincial socioeconomic (eg, gross domestic product [GDP] and gender ratio) and epidemiological (eg, HIV incidence) factors associated with the global prevalence of these digital health attitudes?
- RQ4: How do the model-estimated relationships of these macro-level correlates exhibit spatial nonstationarity, and which specific regional contexts catalyze or inhibit the digital manifestation of HIV stigma?

By bridging computational linguistics with behavioral geography, this study provides insights into how digital technology and local environments intersect to shape human reactions to one of the world’s most stigmatized health conditions. Our findings contribute to the vision of “zero discrimination” advocated by the Joint United Nations Programme on HIV/AIDS (UNAIDS) and offer an approach for future research on computer-mediated social support and digital health governance.

### Theoretical Framework: Sociohealth Correlates of Place

In health geography, “place” is conceptualized as a complex intersection of social, demographic, and epidemiological contexts that collectively modulate public health perceptions. We draw on the socioecological model to organize our analytical framework. The socioecological model posits that health-related behaviors and attitudes are shaped by multiple interacting levels: intrapersonal, interpersonal, community, and societal [20]. While our data do not permit individual-level analysis, we apply the socioecological model’s core insight at the macro-level by examining how provincial-level contextual factors, including demographic, epidemiological, and socioeconomic conditions, are associated with regional patterns in digital discourse. Specifically, the socioecological model guides our selection of variables by identifying

3 contextual domains that prior literature has linked to health-related attitudes: population composition (demographic substrate), disease burden (epidemiological context), and material conditions (socioeconomic foundation).

### ***The Demographic and Cultural Substrate***

The demographic environment serves as the cultural and social substrate for empathy and moral judgment, as it defines the normative background of a community [21]. Previous research has established that social environments influenced by different demographic structures can exhibit different health-related communication styles [22-24]. Within the socioecological model, this dimension corresponds to the community-level normative environment that shapes interpersonal attitudes. Gender ratio reflects gendered socialization patterns within a province, which are associated with prevailing moral norms and communication styles regarding sensitive health topics. Life expectancy serves as a proxy for overall health resilience and social stability, indicating the public's baseline health security. Average schooling years, as a primary determinant of health literacy, acts as a cognitive resource that may reduce the proliferation of irrational stigma [23].

### ***The Epidemiological Context***

The epidemiological context draws on the logic of contact theory, which suggests that the visibility and prevalence of a disease within a place may shape public attitudes through increased salience and familiarity [25,26]. We do not equate provincial HIV incidence with direct interpersonal contact; rather, we conceptualize these indicators as structural background factors that condition the discursive environment. Within the socioecological model, this dimension represents the community-level disease burden that shapes risk perception and threat appraisal. HIV incidence rate reflects the current risk and epidemiological salience within a province; high incidence may be associated with either increased social awareness or defensive stigma as a psychological distancing mechanism. HIV mortality rate captures the perceived severity of the disease, which often sustains fear-based stigmas and may inhibit humanitarian support.

### ***The Socioeconomic Foundation***

The socioeconomic foundation provides the material basis for digital literacy and the diversification of values in the digital

public sphere. Economic development influences whether digital discourse is dominated by survival-based anxieties or postmaterialist empathy [27,28]. Within the socioecological model, this dimension captures the community-level resource environment that enables or constrains health-related communication. GDP per capita, as a measure of economic development, is associated with greater exposure to pluralistic health perspectives and higher social capital. Consumption expenditure per capita, representing actual living standards, determines the disposable social and cognitive resources available for digital engagement.

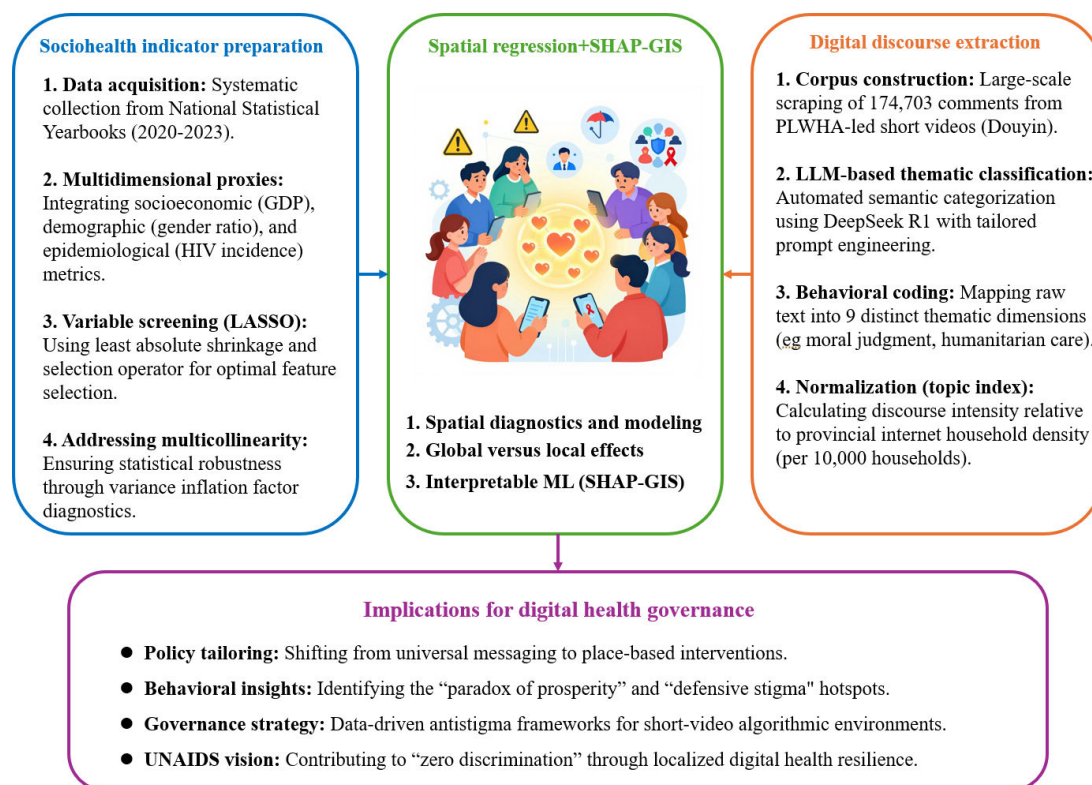
## **Methods**

### ***Overview***

This section delineates the integrated methodological framework designed to investigate the sociospatial correlates of digital HIV/AIDS discourse in China. This study used a cross-sectional ecological design combining provincial-level socioeconomic and epidemiological data with social media comment data. As illustrated in Figure 1, our approach follows a symmetric convergence model that bridges macro-environmental indicators with micro-level digital behaviors. The methodology is structured into 3 primary phases:

1. Sociohealth data and corpus preparation: We systematically acquired provincial socioeconomic and epidemiological indicators and constructed a large-scale digital corpus of 174,394 comments from Douyin.
2. LLM-based behavioral classification: Using the DeepSeek R1 large language model, we categorized the discourse into 9 thematic dimensions to quantify public perceptions across mainland China.
3. Spatial and interpretable machine learning analysis: We integrated spatial econometric modeling to identify global associations between macro-level contextual factors and digital discourse outcomes and used SHAP-geographic information system (GIS) to decompose these associations into localized marginal contributions, thereby accounting for spatial nonstationarity.

**Figure 1.** An integrated computational pipeline for decoding digital HIV/AIDS discourse. GDP: gross domestic product; GIS: geographic information system; LASSO: least absolute shrinkage and selection operator; LLM: large language model; ML: machine learning; SHAP: Shapley Additive Explanations; UNAIDS: Joint United Nations Programme on HIV/AIDS.



Subsequent subsections provide detailed specifications for the technical implementation and validation of each component.

## Data Preparation

The data collection process began on May 18, 2025, with a keyword search on Douyin using the term “抗艾日常” (roughly translated as “my anti-AIDS life” or “life with HIV/AIDS”). This keyword was selected following a pilot search (May 10–15, 2025), which identified it as the most active hashtag for daily-life narratives of people living with HIV/AIDS on Douyin, capturing first-person content directly relevant to our research question. Importantly, Douyin uses a fuzzy search mechanism for content retrieval, meaning that a single keyword query captures videos containing variations or related expressions of the hashtag, ensuring comprehensive coverage within the thematic domain. The search initially yielded 311 video results. Two independent reviewers screened the titles and descriptions of all 311 videos. The inclusion criteria were (1) people living with HIV/AIDS sharing daily-life experiences, (2) more than 1000 comments,

and (3) Chinese-language content. Interreviewer agreement was 94.5% (Cohen  $\kappa=0.89$ ), with disagreements resolved through consultation with a third reviewer. We acknowledge that this keyword selection biases our sample toward users who actively engage with daily-life content shared by people living with HIV/AIDS, potentially excluding those who engage only with news-based or educational HIV/AIDS content; this is discussed as a limitation in the *Discussion* section.

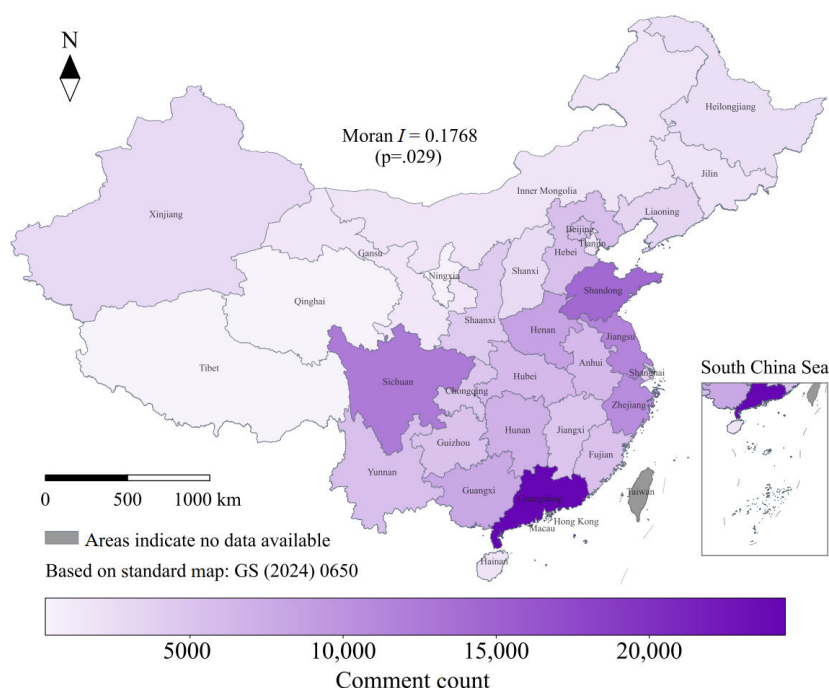
After screening, 91 videos were identified as highly relevant to the research focus (a representative video and typical comments are shown in Figure 2), specifically those in which people living with HIV/AIDS shared their daily life experiences and were also relatively popular (ie, having more than 1000 comments). Subsequently, web scraping techniques were used to collect all publicly available comments under these 91 selected videos, resulting in a dataset comprising 291,264 individual comments.

**Figure 2.** Representative videos and typical comments.

Next, the initial comment data underwent a rigorous cleaning process [29,30]. This multistep procedure included (1) converting emojis into their corresponding Chinese character descriptions (eg, “[Smile]” for a smiling emoji) to facilitate textual analysis; (2) removing all instances of “@user-name” mentions; (3) deleting other nonessential characters, such as special symbols or excessive punctuation that did not contribute to the comment’s meaning; (4) filtering out comments containing fewer than 6 Chinese characters, as they were deemed to possess insufficient informational content for meaningful analysis; and (5) excluding comments originating from IP addresses outside mainland China to focus the

analysis on perceptions within the Chinese sociocultural context. Following these data preparation steps, the final dataset consisted of 174,394 valid comments for the analysis. The specific distribution of comment data is shown in Figure 3. All exclusion criteria were predefined based on observable comment characteristics rather than outcome-related variables.

All 31 provincial-level administrative regions in mainland China were included. For social media data, all publicly available comments under the selected videos during the study period were collected without additional sampling.

**Figure 3.** Distribution of raw comments across 31 provincial-level units in mainland China. All maps are based on the standard map with approval GS (2024) 0650 from the Ministry of Natural Resources of China. Darker shades indicate higher prevalence, while lighter shades indicate lower prevalence. The local Moran  $I$  statistic indicates significant spatial clustering.

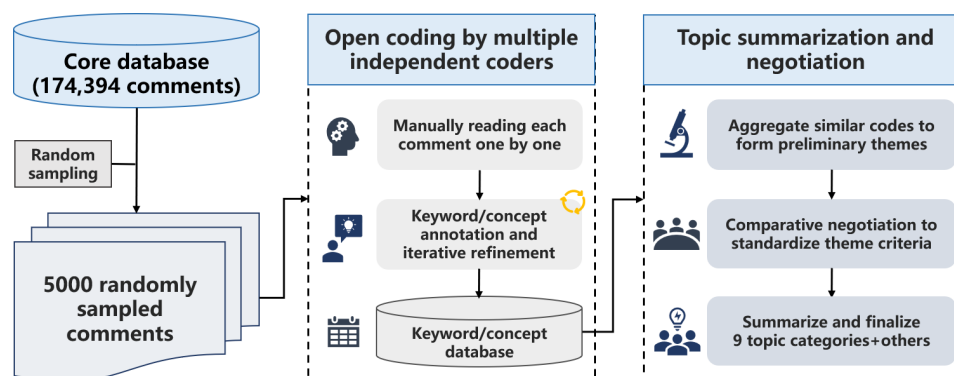
## Text Classification

To comprehensively analyze the rich discussion features within the comments and categorize the diverse public perceptions, we adopted an analytical approach based on LLM-driven few-shot learning. Existing research has demonstrated the high accuracy and efficacy of LLM-driven few-shot learning paradigms for nuanced thematic classification of large-scale social media text data [31,32]. Given these established capabilities, we selected the *DeepSeek R1* API to facilitate the thematic analysis for this study [33]. This

approach was deemed most suitable for systematically and accurately processing our extensive and complex dataset of user comments [34]. The prompt and core Python code used in this study can be found on GitHub [35].

The development of the thematic category framework was initiated by randomly selecting 5000 comments from the cleaned dataset. Adhering to open coding principles [36], the research team iteratively analyzed this subset to identify potential themes. The specific process for identifying the topic is shown in Figure 4.

**Figure 4.** Workflow for identifying topics in accordance with open coding principles.



Through systematic analysis and synthesis, this study identified 10 main thematic categories, each with corresponding definitions, as follows:

1. *Disease knowledge and transmission discussion*: This category focuses on medical knowledge of HIV/AIDS, including transmission patterns, incubation period, symptoms, diagnostic methods, prevention measures, and treatment advances. Comments in this category are characterized by inquiries, information sharing, popular science education, or myth-busting. For example, some users asked, "How do I know if I have AIDS?"
2. *Humor, banter, and ironic expression*: Comments in this category use humor, banter, exaggeration, satire, or playful language to address HIV/AIDS-related topics. Such expressions may serve as a means of emotional regulation but may also contain sarcasm or inappropriate connotations. For example, "Even AIDS can be cured. When will my gout be cured?"
3. *Information sources and credibility discussion*: This category focuses on the accuracy and reliability of medical information, news reports, and expert opinions. Comments typically involve assessments or doubts about the authenticity, authority, and timeliness of the information. For example, "Even the world's top medical teams can't handle this; you're overthinking it."
4. *Moral judgment and attribution of blame*: Comments in this category involve a moral assessment of the causes of infection or patient behavior, often stigmatizing the disease by attributing it to specific behaviors or groups and expressing condemnation, discrimination, or prejudice. For example, "Except for those infected by blood transfusion, others do not deserve sympathy."
5. *Personal experience and current situation sharing*: This category includes users sharing their own infection-related experiences, symptoms, treatment processes, psychological states, or describing individuals' health status and life details after infection. For example, "Just treat it as a chronic disease, take medication on time. Now my viral load is undetectable, and I'm working hard to provide a better life for my family."
6. *Risk underestimation and cognitive bias*: Comments in this category reflect an underestimation of the risk of HIV infection or the severity of the disease or display misconceptions about viral transmission. For example, "Compared to cancer and cardiovascular diseases, AIDS is not a disease."
7. *Social impact and public issue discussion*: This category addresses the association of HIV/AIDS at both individual and societal levels, including social relationships, family, employment, mental health, discrimination, policy recommendations, public health, and legal responsibilities. For example, "Can this kind of situation be prosecuted and sentenced?"
8. *Support, encouragement, and humanitarian care*: Comments in this category express sympathy, understanding, support, and encouragement for people living with HIV/AIDS, call for the elimination of discrimination, respect for privacy, or share positive coping attitudes. For example, "Child, don't listen to what they say. The words of others are meaningless. If you are sick, get treatment and take care of yourself."
9. *Worry, fear, and anxiety emotions*: This category reflects users' concerns and anxieties about infection risks, the disease itself, or social discrimination. For example, "Now I think it's better to stay single, I'm afraid of falling in love."

10. *Others*: This category includes comments that cannot be classified into any of the aforementioned categories, such as content unrelated to HIV/AIDS and people living with HIV/AIDS, meaningless characters, or scattered remarks with indeterminate intent. For example, “This can’t be downloaded, why?”

These predefined themes were then used to guide the DeepSeek R1 671B model in conducting a few-shot classification of the entire 174,394 comments. The model was deployed using the open-source version on our university’s high-performance computing center, without any modification or quantization. The few-shot prompts were designed with clear task definitions, category descriptions accompanied by representative examples, and explicit single-label classification instructions to ensure consistency. All classification prompts and instructions were provided in Chinese to align with the language of the comments. For each comment, the model was instructed to assign exactly one best-fitting category from the predefined theme list, with no option for multiple assignments or uncertain classification. Comments that could not be confidently assigned to any

substantive theme were classified into a residual “Others” category, which was treated as noise and excluded from further thematic analysis. To ensure reproducibility, we set the model temperature to 0. The full classification prompt and core code are provided on GitHub [35].

To validate the LLM’s classification accuracy, a stratified random sample was constructed by selecting 100 comments per category (1000 total) from the classified corpus, ensuring adequate representation of lower-frequency themes. This validation sample was manually coded by 2 independent coders blind to the LLM labels, with intercoder agreement. Disagreements were resolved through discussion. The LLM’s classification for this subset was compared against human-coded labels. The model achieved 95.20% overall accuracy (48/1000 errors) with a Cohen  $\kappa$  of 0.9467, indicating excellent agreement beyond chance [37]. Each thematic category accuracy’s is detailed in Table 1, and the full confusion matrix is provided in Figure SA-1 in Multimedia Appendix 2. These results confirm that the LLM-based classification effectively covers the breadth and complexity of the discussion content in the comments.

**Table 1.** Classification performance assessment based on stratified sampling (100 per category).

| Topic categories                                  | Precision <sup>a,b</sup> | Recall <sup>a,b</sup> | <i>F</i> <sub>1</sub> -score <sup>a,b</sup> |
|---------------------------------------------------|--------------------------|-----------------------|---------------------------------------------|
| Disease knowledge and transmission discussion     | 0.970                    | 0.990                 | 0.980                                       |
| Humor, banter, and ironic expression              | 0.980                    | 0.778                 | 0.867                                       |
| Information sources and credibility discussion    | 0.960                    | 1.000                 | 0.980                                       |
| Moral judgment and attribution of blame           | 0.930                    | 0.959                 | 0.944                                       |
| Personal experience and current situation sharing | 0.900                    | 0.918                 | 0.909                                       |
| Risk underestimation and cognitive bias           | 0.980                    | 1.000                 | 0.990                                       |
| Social impact and public issue discussion         | 0.940                    | 0.979                 | 0.959                                       |
| Support, encouragement, and humanitarian care     | 0.990                    | 0.990                 | 0.990                                       |
| Worry, fear, and anxiety emotions                 | 0.980                    | 1.000                 | 0.990                                       |
| Others                                            | 0.890                    | 0.957                 | 0.922                                       |

<sup>a</sup>Macro average: precision=0.952; recall=0.957; *F*<sub>1</sub>-score=0.953.  
<sup>b</sup>Weighted average: precision=0.953; recall=0.952; *F*<sub>1</sub>-score=0.951.

Influencing Factor Analysis

To analyze the socioeconomic correlates of localized HIV/AIDS discourse, we implemented a multistage quantitative framework. This approach transitioned from variable selection and global regression to spatial diagnostics and advanced local interpretation.

Dependent Variable: The Normalized Topic Index

The primary dependent variable in this study is the Normalized Topic Index, representing the intensity of public engagement with specific HIV/AIDS-related themes across the 31 provinces. To ensure that our analysis reflects regional social attitudes rather than mere differences in digital infrastructure or population size, we normalized the absolute discussion counts (ie, comment volume) for each province [38]. The normalization was performed by dividing the total comment count for a specific topic by the number of mobile internet access users in that province:

$$Topic\ index_i = \frac{Comment\ count_i}{Mobile\ internet\ access\ users_i}$$

This procedure effectively controls for digital penetration bias, allowing for a more equitable comparison of “place-based” health perceptions across regions with varying levels of internet accessibility. Mobile internet access users were selected because they represent the population with the technical capacity to engage with mobile short-video platforms. This denominator does not capture platform-specific user demographics (eg, Douyin’s age and urban-rural composition), which may introduce residual bias.

Feature Selection via Least Absolute Shrinkage and Selection Operator Regression

On the basis of the theoretical framework established in section 2, we gathered a comprehensive dataset of 8 candidate indicators to represent the demographic, epidemiological,

and socioeconomic characteristics of each province. The data were primarily sourced from the National Bureau of Statistics of China and the China Health Statistical Yearbook. To ensure comparability across variables with different units and scales, all independent variables were standardized. The statistical distribution and regional disparities of these standardized candidate variables ( $V_1$ - $V_8$ ) are visualized in Figure SA-2 in [Multimedia Appendix 3](#). The violin plots reveal significant spatial heterogeneity across the 31 provinces, particularly in economic output ( $V_1$ ) and HIV epidemiological metrics ( $V_7$  and  $V_8$ ), providing the necessary empirical variance for spatial analysis.

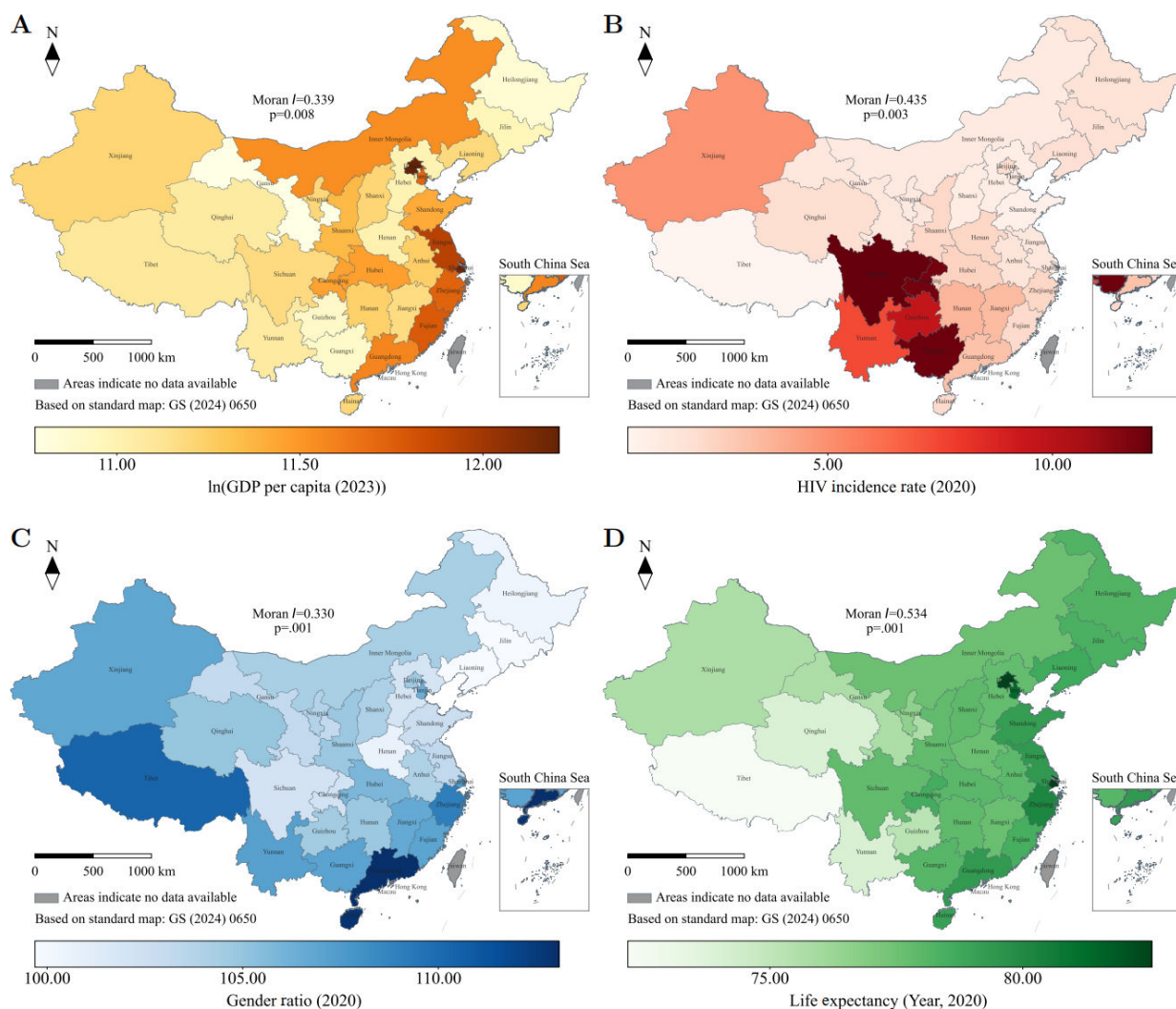
However, a significant methodological challenge arises from the limited sample size of 31 provincial units. Including all 8 theoretical candidates in the spatial econometric models would severely deplete the degrees of freedom and introduce a substantial risk of overfitting, which could lead to unstable and unreliable parameter estimates. To address this, we used least absolute shrinkage and selection operator (LASSO) regression to perform data-driven feature selection [39]. LASSO was implemented using scikit-learn version 1.8.0 in Python. The regularization parameter  $\lambda$  was selected via 5-fold cross-validation, using  $\lambda_{\min}$  (the value minimizing mean squared error). The 1-SE rule (selecting the largest  $\lambda$  within 1 SE of  $\lambda_{\min}$ ) yielded identical selections. By introducing an L1 penalty term, LASSO identifies the most resilient predictors while shrinking the coefficients

of redundant or less contributory variables to zero. We conducted multiple iterations of the selection process to evaluate the stability of each variable. As illustrated in Figure SA-3 in [Multimedia Appendix 4](#), four variables demonstrated superior selection frequency.

To ensure the statistical validity of the subsequent modeling, we performed preliminary diagnostics focusing on multicollinearity and spatial dependence. First, a multicollinearity check was conducted for the 4 selected independent variables [40]. As shown in Figure SA-4 in [Multimedia Appendix 5](#), all values remained well below the conservative threshold of 5, indicating that multicollinearity does not pose a threat to the stability of our parameter estimates.

Second, we used GIS-based multipanel mapping to visualize the geographic distribution of these indicators across the 31 provinces. To quantify the observed spatial patterns, we calculated the global Moran  $I$  index [41]. As labeled in the composite maps ([Figure 5](#)), all 4 variables exhibited significant positive spatial autocorrelation. The consistently significant and positive Moran  $I$  values confirm a high degree of spatial clustering across China's provincial landscape. These findings provide a definitive statistical justification for transitioning from a nonspatial ordinary least squares (OLS) framework to spatial econometric models (spatial lag model [SLM]/spatial error model [SEM]) to account for spatial dependence and spillover estimated associations.

**Figure 5.** Spatial distribution characteristics of the 4 selected variables across 31 provincial-level units in mainland China. Darker shades indicate higher prevalence, while lighter shades indicate lower prevalence. The global Moran  $I$  statistic indicates significant spatial clustering.



## Regression Modeling and Robustness

We first estimated an OLS-based multiple linear regression (MLR) model for each thematic outcome to establish a nonspatial benchmark [42]. Model diagnostics were then conducted to evaluate whether this benchmark specification was adequate. Specifically, we applied the Ramsey RESET test to examine potential functional-form misspecification and the Breusch-Pagan test to assess heteroscedasticity [43]. When heteroscedasticity was detected, robust SEs were used to improve the reliability of statistical inference. In addition, we calculated Moran  $I$  for the OLS residuals to determine whether spatial dependence remained after accounting for the selected socioeconomic and epidemiological predictors.

For topic models exhibiting significant residual spatial autocorrelation, we further implemented spatial econometric models following a general-to-specific strategy [44]. We constructed a queen-contiguity spatial weight matrix  $W$  for the 31 provinces using the `Queen.from_dataframe` method from the `libpysal` library, in which provinces sharing a common border or vertex were defined as neighbors. The matrix was not row standardized. As a robustness check, we also used an inverse-distance weight matrix ( $w_{ij} = 1/d_{ij}$ )

based on provincial centroid distances, and the results were substantively consistent.

The spatial analysis began with the spatial Durbin model (SDM), which accounts for spatial lags in both the dependent variable and the independent variables. We then used Wald and likelihood ratio tests to examine whether the SDM could be simplified into an SLM or an SEM. This procedure allowed us to select a parsimonious spatial specification that captured the relevant form of spatial dependence while avoiding unnecessary overparameterization.

Finally, to control for the inflation of type I errors arising from multiple topic-specific analyses, we applied the false discovery rate (FDR) correction [45]. All  $P$  values from the final selected models were adjusted using the Benjamini-Hochberg procedure, and associations that remained significant at the FDR-adjusted 0.05 level were reported as robust findings. Given the cross-sectional ecological design of this study, these results should not be interpreted as causal effects. Rather, they characterize regional-level associations between macrocontextual conditions and digital HIV/AIDS discourse and should not be used to infer individual-level psychological mechanisms.

## Interpretation of Localized Contributions via SHAP-GIS

While the MLR and spatial econometric models described earlier provide a robust estimation of “global” average associations across the country, they may mask spatial nonstationarity where the influence of a socioeconomic driver varies significantly from one province to another. To uncover these geographical nuances, we used the SHAP framework to decompose the model’s predictions [46].

Unlike traditional regression coefficients that offer a single aggregate value, the SHAP approach quantifies the marginal contribution of each independent variable for every specific provincial observation [47]. By integrating these province-specific SHAP values with GIS, we generated localized effect maps [48]. This SHAP-GIS coupling allows us to visualize the spatial heterogeneity of the correlates, identifying where and why certain socioeconomic factors (eg, GDP or gender ratio) act as factors associated with increased stigmatizing discourse or inhibitors of specific HIV/AIDS discourses [49]. This transition from global averages to localized interpretation provides the nuanced empirical evidence required for place-based public health interventions. SHAP-GIS was applied only to OLS models that did not exhibit significant spatial autocorrelation (tested via Moran *I* on residuals). For models with significant spatial dependence, we did not apply SHAP-GIS and instead relied on spatial lag and spatial error models for interpretation.

## Ethical Considerations

This study was conducted in accordance with the ethical standards of the responsible committee on human

experimentation and with the Declaration of Helsinki. The research protocol, entitled “Exploring Human Behavior in the Social Media Landscape” (principal investigator: PH), was reviewed and approved by the institutional review board of the Hong Kong University of Science and Technology (Guangzhou). The ethics approval reference number was HKUST(GZ)-HSP-2024-0065. This study exclusively involved the analysis of publicly available social media data. No data were collected directly from human participants, and no identifiable personal information was used or stored during this research. All data handling procedures were performed in strict compliance with the platform’s terms of service and applicable privacy regulations to ensure the confidentiality and anonymity of the data sources.

## Results

### Descriptive Results: The Landscape of Digital Discourse

The thematic classification of the 174,394 collected comments revealed a multifaceted and somewhat polarized public reaction to HIV/AIDS narratives. As summarized in Table 2, the discourse was characterized by a tension between traditional moralizing perspectives and emerging humanitarian support. The most prevalent identifiable theme was moral judgment and attribution of blame (41,663/174,394, 23.89%). This high proportion indicated that HIV/AIDS remains heavily framed through a moral lens in the Chinese digital sphere, with a significant segment of the audience focusing on the “responsibility” or “fault” of the individuals portrayed in the videos.

Table 2. Classification results, raw and weighted considered.

| Topic name                                        | Count, n (%)   |
|---------------------------------------------------|----------------|
| Moral judgment and attribution of blame           | 41,663 (23.89) |
| Disease knowledge and transmission discussion     | 23,344 (13.39) |
| Support, encouragement, and humanitarian care     | 18,101 (10.38) |
| Social impact and public issue discussion         | 14,908 (8.55)  |
| Humor, banter, and ironic expression              | 14,580 (8.36)  |
| Risk underestimation and cognitive bias           | 9326 (5.35)    |
| Worry, fear, and anxiety emotions                 | 7693 (4.41)    |
| Personal experience and current situation sharing | 4932 (2.83)    |
| Information sources and credibility discussion    | 4615 (2.65)    |
| Others                                            | 35,232 (20.2)  |

While disease knowledge and transmission discussion (23,344/174,394, 13.39%) reflected a substantial public desire for medical facts, the presence of risk underestimation and cognitive bias (9326/174,394, 5.35%) suggested that misinformation and a lack of perceived vulnerability persisted among users. Encouragingly, support, encouragement, and humanitarian care (18,101/174,394, 10.38%) emerged as a major pillar of discourse. This suggested that social media also acted as a vital space for prosocial engagement and the expression of empathy toward people living with HIV/AIDS. Emotional expressions ranging from worry and fear

(7693/174,394, 4.41%) to humor and banter (14,580/174,394, 8.36%) reflected the complex psychological landscape of the audience. The “Others” category (35,232/174,394, 20.20%) highlighted the vast amount of miscellaneous interaction common on short-form video platforms. A post hoc analysis of 200 randomly sampled comments from this category revealed that the majority comprised platform-specific interactions (83/200, 41.5%), off-topic discussions (56/200, 28.0%), fragmented or ambiguous text (35/200, 17.5%), and comments about video production (26/200, 13%), suggesting that this residual category primarily reflects the diverse

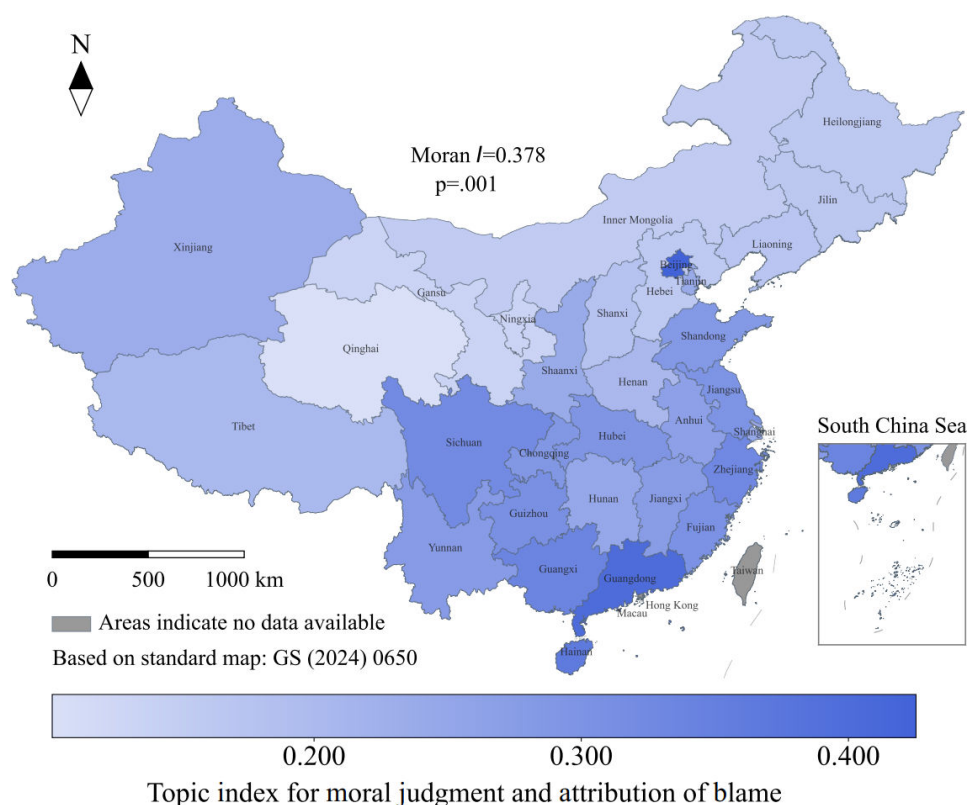
interactional nature of short-video discourse rather than undetected stigmatizing content.

## Spatial Distribution of HIV/AIDS Digital Discourse

To uncover the geographic footprint of public health perceptions, we mapped the Topic Index across the 31 provinces. While the full provincial dataset is provided in Figure SA-5 in [Multimedia Appendix 6](#), this section focuses on the spatial patterns of the most prominent themes: moral judgment, support, and disease knowledge. Normalization revealed significant spatial heterogeneity, indicating that a province's socioeconomic and cultural context deeply influenced the nature of its digital discourse.

As the most prevalent theme, moral judgment exhibited the highest intensity across China; yet, with clear regional disparities. As shown in [Figure 6](#), the highest engagement levels were concentrated in major political and economic centers, notably Beijing (0.4253) and Guangdong (0.3991). These regions acted as provinces with high discursive intensity, where narratives were most likely to be scrutinized through a moral lens. High indices were also observed in Guangxi (0.3370) and Zhejiang (0.3261). Conversely, inland provinces such as Gansu (0.1347) and Qinghai (0.1020) showed significantly lower intensity, suggesting a different cultural or digital framing of HIV narratives in these “places.”

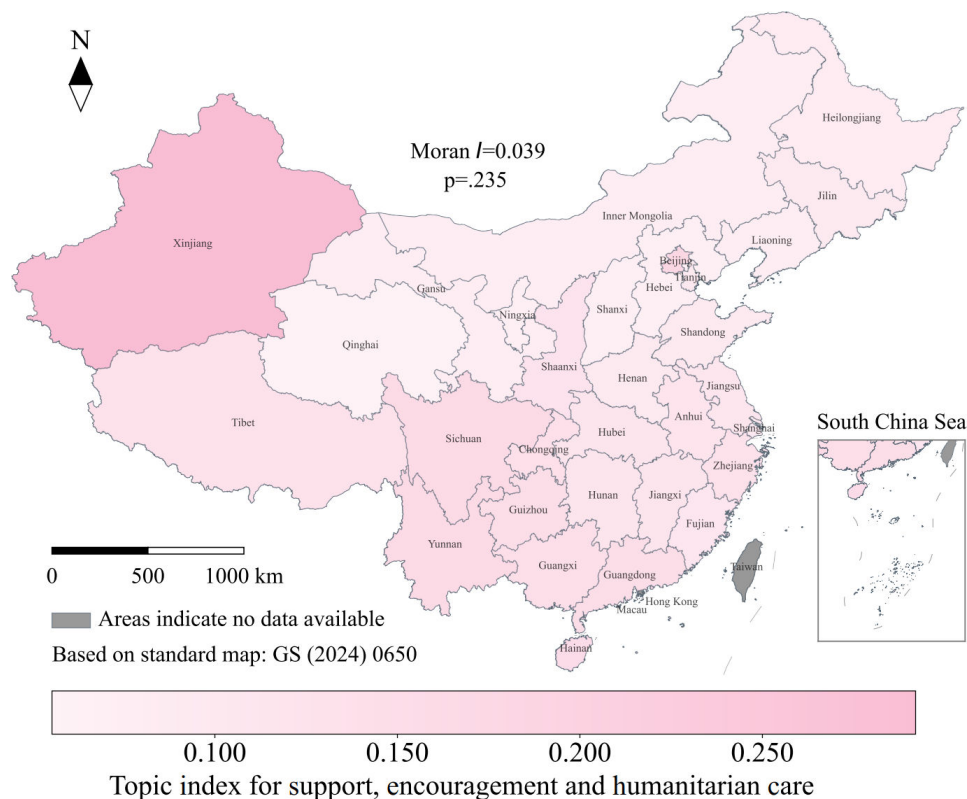
**Figure 6.** Spatial distribution for the Topic Index of moral judgment and attribution of blame across 31 provincial-level units in mainland China (excluding Taiwan, Hong Kong, and Macao). The index represents the prevalence of moral judgment and attribution of blame comments, calculated as the frequency of comments classified into this theme per 10,000 mobile internet users at the provincial level. Darker shades indicate higher prevalence, while lighter shades indicate lower prevalence. The global Moran  $I$  statistic indicates significant spatial clustering.



The distribution of supportive social discourse exhibited a distinctive geographical pattern and did not strictly follow economic logic. As shown in [Figure 7](#), Xinjiang stood out with the highest support index (0.2920) in the country. This finding was noteworthy and may reflect a confluence of factors, including region-specific public health messaging, distinct cultural norms regarding health discourse, or platform moderation practices. Further investigation with

contextual data is needed to identify the specific mechanisms. Beijing (0.1733) and Yunnan (0.1665) also showed high levels of humanitarian care, reflecting the coexistence of supportive and stigmatizing discourses within the same geographic boundaries. Despite their economic similarities, Jiangsu (0.1102) and Shandong (0.0910) exhibited lower support indices than Guangdong, their southern neighbor.

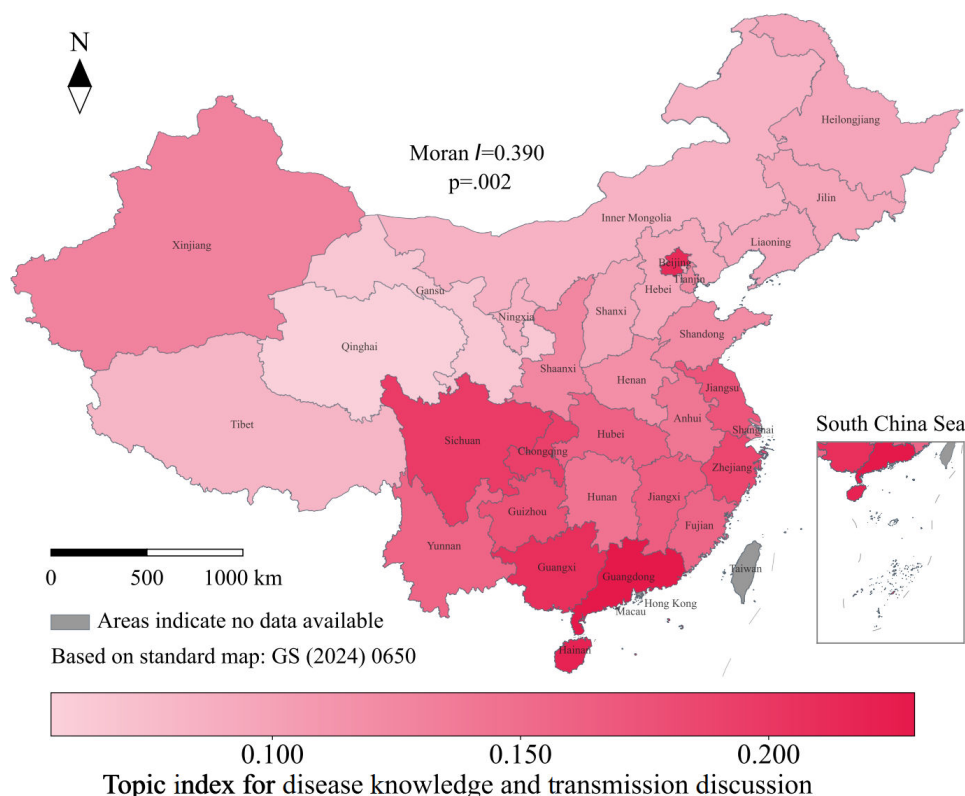
**Figure 7.** Spatial distribution for Topic Index of support, encouragement, and humanitarian care across 31 provincial-level units in mainland China (excluding Taiwan, Hong Kong, and Macao). The index represents the prevalence of support, encouragement, and humanitarian care comments, calculated as the frequency of comments classified into this theme per 10,000 mobile internet users at the provincial level. Darker shades indicate higher prevalence, while lighter shades indicate lower prevalence. The global Moran  $I$  statistic does not indicate significant spatial clustering.



Disease knowledge and transmission discussion showed a spatially uneven pattern shaped by both health information capacity and local epidemiological salience. As shown in Figure 8, developed provinces and major digital hubs, including Guangdong (0.2294), Beijing (0.2137), Zhejiang (0.1856), and Jiangsu (0.1717), exhibited relatively high Topic Indices, possibly reflecting stronger health literacy, information access, and digital engagement. Meanwhile, several provinces with higher HIV epidemiological salience,

such as Guangxi (0.2075), Sichuan (0.1960), Chongqing (0.1923), Yunnan (0.1578), and Xinjiang (0.1293), also showed elevated levels of knowledge-oriented discussion, suggesting that public information demand may be stronger where HIV/AIDS is more visible. In contrast, Qinghai (0.0554) and Gansu (0.0654) reported the lowest intensities, indicating weaker normalized engagement with medical and transmission-related HIV/AIDS information.

**Figure 8.** Spatial distribution of disease knowledge and transmission discussion across 31 provincial-level units in mainland China (excluding Taiwan, Hong Kong, and Macao). The index represents the prevalence of disease knowledge and transmission discussion comments, calculated as the frequency of comments classified into this theme per 10,000 mobile internet users at the provincial level. Darker shades indicate higher prevalence, while lighter shades indicate lower prevalence. The global Moran  $I$  statistic indicates significant spatial clustering.



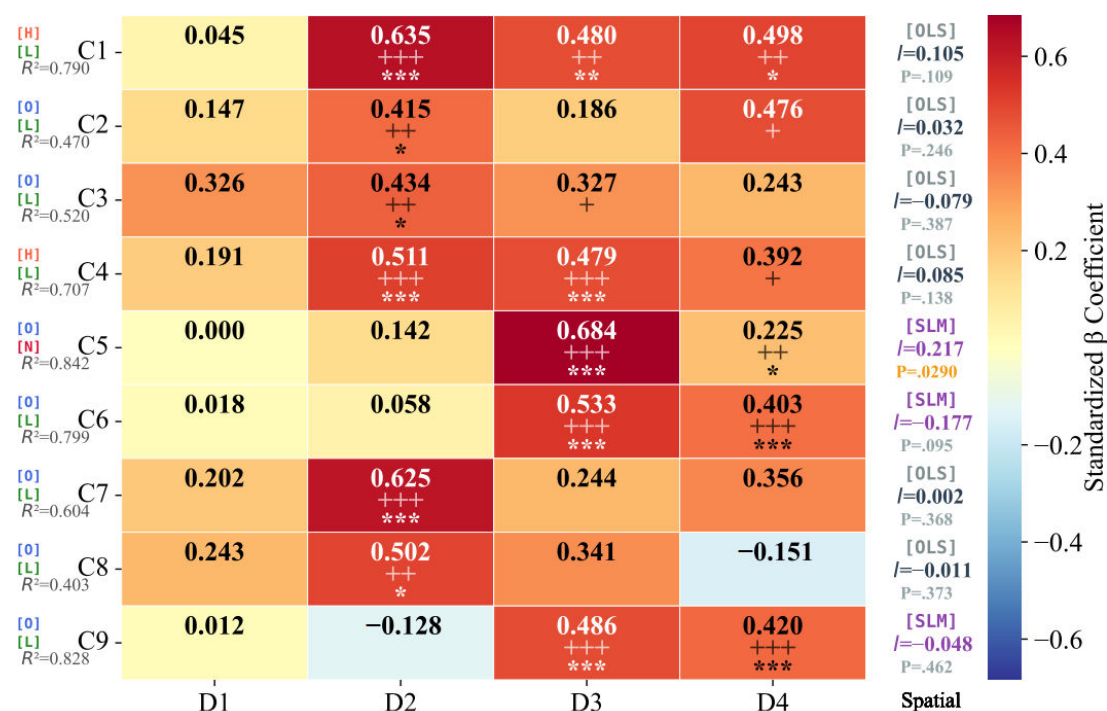
## Regression Analysis: Sociohealth Correlates of Digital Discourse

### Overview

To systematically identify the correlates of provincial discourse, we estimated regression models for 9 thematic topics. The core results, integrated with spatial diagnostics and multiple-testing corrections, are synthesized in the FDR-corrected standardized  $\beta$  heatmap (Figure 9). Owing to the large volume of regression outputs (9 topic models  $\times$  multiple predictors), full results, including all coefficients, SEs, CIs,  $P$  values, and model fit statistics, are provided on GitHub [35]. The models achieved substantial explanatory

power, with  $R^2$  (McFadden  $R^2$  for the SLM) values ranging from 0.403 (support) to 0.842 (personal experience). Diagnostic tests identified significant spatial autocorrelation in the residuals for personal experience, risk underestimation, and worry and anxiety. Consequently, an SLM was used for these topics (labeled as [Lag] in Figure 9) to account for spatial dependence and potential spillover patterns among neighboring provinces. By applying the FDR correction, we ensured that the reported associations were robust against type I errors. Furthermore, for topics such as moral judgment where heteroscedasticity was detected ([H]), robust SEs were used to ensure the validity of our inferences.

**Figure 9.** Predictive effects of socioeconomic and health indicators (D1-D4) on HIV/AIDS–related social media discussion topics (C1-C9). The matrix displays standardized regression coefficients with false discovery rate (FDR)–corrected significance levels (\*, \*\*, \*\*\*), heteroscedasticity tests ([O]/[H]), and linearity tests ([L]/[N]). The spatial analysis column reports residual Moran *I* values and the final spatial regression approach (eg, SLM). GDP: gross domestic product; OLS: ordinary least squares; SLM: spatial lag model.



FDR-corrected significance ( $\alpha=0.05$ ): \*  $q<0.05$ , \*\*  $q<0.01$ , \*\*\*  $q<0.001$

Original significance: +  $p<0.05$ , ++  $p<0.01$ , +++  $p<0.001$

Ramsey RESET Test: [L]=Linearity supported ( $p\geq 0.05$ ), [N]=Potential nonlinearity ( $p<0.05$ )

Breusch-Pagan Test: [O]=Homoskedasticity ( $p\geq 0.05$ ), [H]=Heteroskedasticity (robust SE used)

Spatial: [XXX] is the used model, *I* and *P* mean residual Moran *I* and *p*-value

Independent variables:

D1: ln(GDP per capita (2023)) D2: HIV incidence rate (2020) D3: Gender ratio (2020)

D4: Life expectancy (2020)

Topic category:

C1: Disease knowledge and transmission discussion C2: Humor, banter and ironic expression

C3: Information sources and credibility discussion

C4: Moral judgment and attribution of blame

C5: Personal experience and current situation sharing

C6: Risk underestimation and cognitive bias C7: Social impact and public issue discussion

C8: Support, encouragement and humanitarian care C9: Worry, fear and anxiety emotions

The HIV incidence rate emerged as the most consistent and powerful driver across multiple discourse dimensions, reinforcing the “contact theory” in a digital context. Higher incidence rates were most strongly associated with disease knowledge ( $\beta=.635$ ;  $q<.001$ ) and social impact discussion ( $\beta=.625$ ;  $q<.001$ ). Crucially, it was the primary factor associated with increased stigmatizing discourse for moral judgment ( $\beta=.511$ ;  $q<.001$ ), suggesting that in provinces with higher disease visibility, the public was more likely to engage in both factual inquiry and moralistic scrutiny. The HIV incidence rate was also positively associated with support, encouragement, and humanitarian care ( $\beta=.502$ ;  $q=.012$ ), indicating that higher epidemiological salience corresponded not only to stronger stigmatizing discourse but also to greater prosocial engagement. This pattern suggested that disease visibility may intensify public attention in multiple directions, producing both moralized scrutiny and humanitarian responses.

The gender ratio representing the demographic substrate of a province showed a unique influence on subjective themes. Provinces with higher male-to-female ratios were significantly more likely to engage with personal experience sharing ( $\beta=.684$ ;  $q<.001$ ) and risk underestimation ( $\beta=.533$ ;  $q<.001$ ). This suggested that the demographic composition of a “place” may shape the communal appetite for anecdotal versus factual content. This also contributed significantly to moral judgment ( $\beta=.479$ ;  $q<.001$ ), hinting that demographic imbalances may correlate with more rigid or judgment-heavy social norms in digital spaces.

As a proxy for public health resilience, higher life expectancy was positively associated with disease knowledge ( $\beta=.498$ ;  $q=.018$ ) and humor and banter ( $\beta=.476$ ;  $P=.03$ ). This suggested that in regions with superior health capital, the public might feel more “empowered” to discuss HIV through either intellectual inquiry or social levity. Surprisingly,

$\ln(\text{GDP per capita})$  showed relatively modest standardized  $\beta$  values compared to epidemiological and demographic factors. Its most notable (although marginally significant) association was on support ( $\beta=.243$ ;  $P=.32$ ), suggesting that economic prosperity provided a baseline for humanitarian expression but did not dictate the emotional tone of the discourse as strongly as the actual disease burden did.

### Localized Correlates of Discourse: SHAP-GIS Interpretation

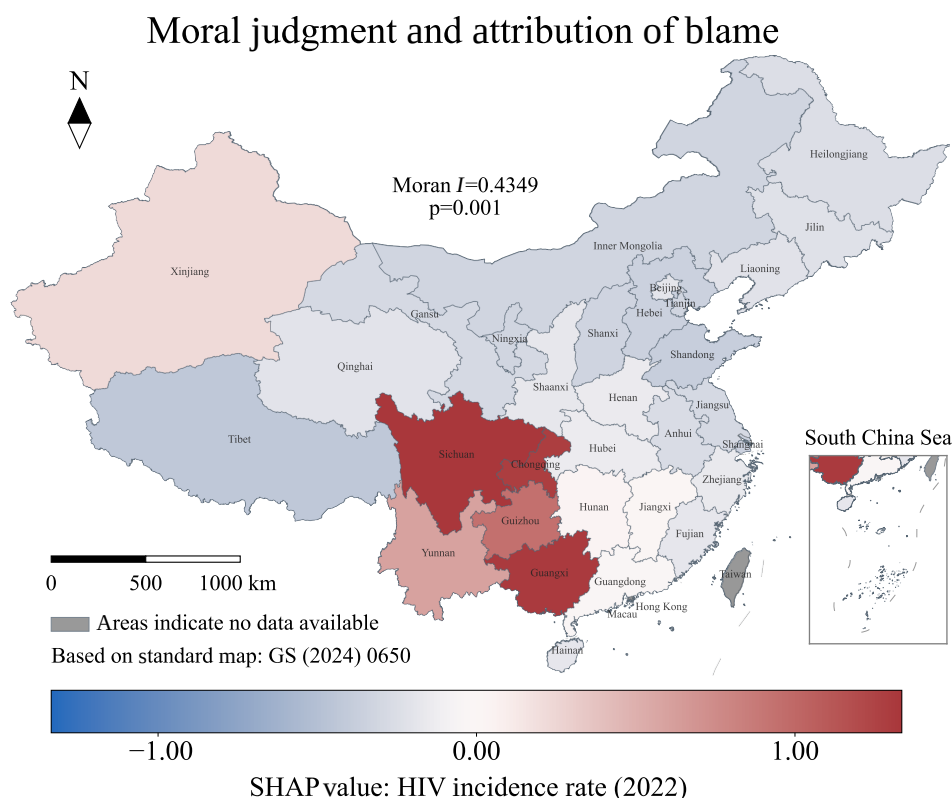
While the previous section established global associations, the global coefficients might mask significant spatial heterogeneity. We applied the SHAP-GIS framework to the OLS models of 3 representative topics to visualize the “local force” exerted by key socioeconomic correlates across different provinces.

### Localized Association of HIV Incidence on Moral Judgment

The SHAP-GIS analysis for moral judgment and attribution of blame (Figure 10) revealed a clear geographical divide

in the localized association between HIV incidence rate and moralizing discourse. While the global standardized  $\beta$  coefficient indicated a strong positive association between HIV incidence and moral judgment ( $\beta=.511$ ;  $q<.001$ ), the localized SHAP values, ranging from  $-0.432$  to  $1.336$ , showed that this association varied substantially across provinces. The highest positive SHAP values were concentrated in Southwest China. Sichuan (SHAP=1.336), Guangxi (SHAP=1.313), Chongqing (SHAP=1.287), and Guizhou (SHAP=0.940) exhibited the largest positive localized contributions of HIV incidence rate to the predicted intensity of moral judgment and attribution of blame. In these provinces, relatively high HIV incidence values, such as Sichuan at 12.21 per 100,000 population, were associated with stronger model-predicted moralizing discourse. This pattern suggested that in high-prevalence provinces, epidemiological salience might coincide with stronger moral attribution in digital HIV/AIDS discussions.

**Figure 10.** Spatial distribution of localized Shapley Additive Explanations (SHAP) values for HIV incidence rate on moral judgment and attribution of blame. The color gradient represents the magnitude of the SHAP values (blue=buffering and red=amplifying), and the global Moran  $I$  statistic indicates significant spatial clustering.



Conversely, in regions with lower epidemiological visibility, HIV incidence rate showed negative localized SHAP values for moral judgment. Tibet ( $-0.432$ ), Shandong ( $-0.384$ ), and Shanghai ( $-0.346$ ) exhibited the most negative SHAP values. In these contexts, relatively low HIV incidence values, such as 0.55 per 100,000 population in Tibet, were associated with lower model-predicted levels of moralizing discourse. This

pattern indicated that lower local epidemiological salience corresponded to weaker moral attribution in the model. Notably, Yunnan, despite its historically high disease burden (7.37 per 100,000 population), showed a more moderate SHAP value (0.602) compared with neighboring provinces such as Sichuan and Guangxi. This relative moderation suggested that the relationship between HIV incidence and

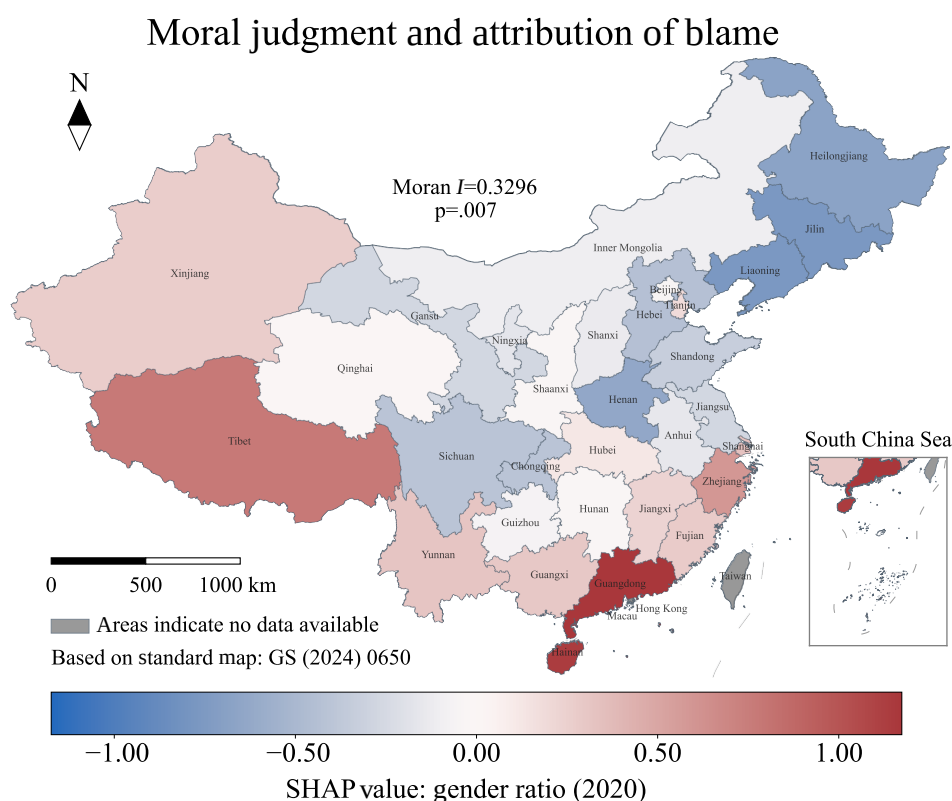
moral judgment might be shaped by additional contextual factors, such as long-term public health communication, local awareness, or region-specific social responses to HIV/AIDS. However, these possible explanations should be interpreted cautiously, as the present cross-sectional ecological design cannot establish causal mechanisms.

### Localized Association of Demographic Substrate on Moral Judgment

As identified in the LASSO selection process, gender ratio was one of the most stable demographic predictors of digital HIV/AIDS discourse. Figure 11 shows that the

localized contribution of gender ratio exhibited a marked regional divide, with SHAP values ranging from  $-0.770$  to  $1.174$ . Provinces with relatively male-skewed demographic structures tended to show positive localized SHAP values for moral judgment and attribution of blame. Guangdong (SHAP= $1.174$ ; feature= $113.08$ ) and Hainan (SHAP= $1.142$ ; feature= $112.86$ ) exhibited the strongest positive localized contributions. Positive SHAP values were also observed in Tibet ( $0.773$ ) and Zhejiang ( $0.587$ ). In these provinces, higher male-to-female ratios were associated with higher model-predicted intensity of moral judgment.

**Figure 11.** Spatial distribution of localized Shapley Additive Explanations (SHAP) values for gender ratio on moral judgment. The color gradient represents the magnitude of the SHAP values (blue=buffering and red=amplifying), and the global Moran  $I$  statistic indicates significant spatial clustering.



Conversely, negative localized SHAP values were observed in Northern China, particularly in the Northeast. Jilin ( $-0.770$ ), Liaoning ( $-0.769$ ), and Heilongjiang ( $-0.675$ ) exhibited the most negative SHAP values in the country. These provinces also had relatively low gender ratios, such as 99.69 in Jilin, indicating a slightly female-skewed or more balanced population structure. In these contexts, a more balanced demographic composition was associated with lower model-predicted moral judgment. These patterns suggested that demographic structure was spatially heterogeneous in its association with digital HIV/AIDS stigma, but further research is needed to clarify the social mechanisms behind this relationship.

## Discussion

### Principal Findings

This study revealed that discussions on Chinese social media platforms about the proactive sharing of daily-life narratives by people living with HIV/AIDS exhibited a complex and multifaceted structure across different regions and socioeconomic contexts. The following interpretations, while grounded in our empirical findings, were offered as hypotheses for future research and should not be considered definitive explanations.

## ***Thematic Landscape of Digital HIV/AIDS Discourse: From Moral Judgment to Supportive Care***

The thematic classification of more than 170,000 comments revealed that the Chinese digital sphere remained a contested terrain for HIV/AIDS discourse. The dominance of moral judgment (23.89%) suggested that traditional stigmatization and “othering” processes remain deeply rooted in public perception. However, the substantial presence of humanitarian support (10.38%) and disease knowledge (13.39%) indicated a growing counter-discourse characterized by health literacy and digital empathy. This pluralistic landscape reflected a transition in public health communication, where short-form video platforms such as Douyin provide a space in which old stigmas and empathy-oriented public discourse coexist [50].

## ***Spatial Heterogeneity of Stigma- and Support-Oriented Discourse***

Our GIS analysis indicated that “place” was associated with the intensity of these digital interactions. The high Topic Index for moral judgment in major hubs such as Beijing and Guangdong suggested that high digital penetration did not inherently correspond to more progressive discourse; rather, it might coincide with stronger expressions of existing social tensions [51]. The high engagement with both knowledge and stigma in provinces such as Guangxi and Sichuan is consistent with the possible relevance of localized disease visibility [16]. The elevated level of support in Xinjiang deviates from national economic trends, suggesting that specific regional cultural narratives, public health campaigns, or viral events may be associated with distinctive discourse environments within the digital landscape. Further research with contextual data is needed to identify the underlying mechanisms.

## ***Sociohealth Correlates in the Geography of Digital HIV/AIDS Stigma***

The global regression analysis highlighted the associated role of HIV incidence and gender ratio in relation to digital discourse. Higher incidence was associated with both knowledge seeking and increased moral judgment [52]. This was consistent with the hypothesis that proximity to the epidemic might be associated with a defensive “risk-distancing” response, where moral attribution may appear alongside heightened disease visibility [53]. We noted that our operationalization of epidemiological proximity, provincial HIV incidence rate, differs from traditional intergroup contact measured at the individual level. We interpreted these ecological associations as evidence that epidemiological salience might be related to discursive environments through visibility and perceived risk, rather than as a direct test of interpersonal contact theory [26]. The selection of gender ratio as the top LASSO predictor reflected the association between population structure and social norms [54]. In male-skewed demographic “places,” the discourse tended to be more rigid and judgment heavy, identifying gender balance

as a crucial, although often overlooked, indicator of a region’s discourse environment.

## ***Gender Balance and Moralized Health Norms***

Our global and localized analyses underscored gender ratio as a consistently strong correlate of HIV/AIDS stigma. The robust SHAP-GIS findings regarding the north-south divide suggested that “place” in digital China is partly associated with its demographic substrate. The higher level of moral judgment in the male-skewed South versus the relatively lower moral judgment intensity in the balanced Northeast indicated that gendered social environments were associated with the “cultural baseline” for empathy and blame. However, we noted that in provinces such as Guangdong, the elevated gender ratio was substantially driven by large floating or migrant populations rather than birth-cohort imbalances alone. The intersection of demographic skew with population mobility and urbanization might be jointly associated with discourse environments. Future studies should disentangle the relative contributions of demographic structure and migration patterns. This suggested gender balance not merely as a demographic statistic, but as a potentially important social factor in health communication that warrants further investigation [19].

## ***Policy Implications for Place-Based Digital Health Governance***

Before discussing specific implications, it is important to clarify what we mean by “stigma” in the context of intervention. In this study, we operationally implement stigma as expressions of social disapproval, blame attribution, and negative stereotyping toward people living with HIV/AIDS in publicly observable online discourse. This reflects both the presence of objective misconceptions among the commenting public and the discursive environment that people living with HIV/AIDS may perceive as stigmatizing. Our data do not distinguish between patients’ perceived stigma and nonpatients’ expressed prejudice, but we treat the 2 as interrelated: expressed stigma in public discourse may contribute to the perceptual environment that shapes patients’ lived experiences.

Our findings identify a “stigma-visibility” association in Southwest China, where elevated HIV incidence is correlated with heightened moral judgment. This pattern may be interpreted through the lens of defensive attribution theory: in high-prevalence provinces such as Sichuan and Guangxi, the epidemiological context may be associated with a psychological distancing pattern [55]. By framing HIV/AIDS through a lens of individual moral culpability, the digital public may symbolically distance the disease from the self, thereby reinforcing an “othering” frame around the disease [56]. As the public in these regions may use moral attribution as a form of perceived-risk distancing [52], digital campaigns should prioritize humanizing the narratives of people living with HIV/AIDS to reduce the “othering” process associated with defensive stigma [57].

The high discursive intensity in tier-1 cities such as Beijing and Shanghai creates a unique challenge: these hubs show both high empathy and high moral scrutiny. Platforms must move beyond simple content moderation. In these polarized “intensity hubs,” algorithms should be calibrated to promote constructive advocacy over moralistic conflict, helping ensure that high digital engagement is associated with social inclusion rather than discursive noise [15]. Policymakers should leverage the high digital literacy in these regions to pilot innovative “buddy programs” and volunteer networks that can bridge traditional social gaps.

The robust association between gender ratio and moral judgment, with positive SHAP contributions clustered in the male-skewed South and West, highlights the relevance of social structure in health communication [58]. In provinces such as Guangdong and Hainan, antistigma campaigns must account for gendered social norms. Messaging should explicitly challenge rigid moral frameworks often observed in male-skewed demographic environments. Encouraging a more diverse range of voices in digital discourse may help buffer against moralized reactions associated with demographic imbalances [59].

## Limitations and Future Work

While this study offered a novel spatial analysis of digital HIV/AIDS discourse, several limitations warrant acknowledgment. First, our analyses were conducted at the provincial level ( $N=31$ ), and the observed associations between provincial characteristics and discourse patterns should not be interpreted as individual-level psychological processes due to the risk of ecological fallacy. Multilevel studies with individual-level data are needed to validate these ecological patterns.

Second, social media data were subject to self-selection bias, as digital discourse is predominantly shaped by a “vocal minority,” potentially excluding the perspectives of the “silent majority.” The keyword “抗艾日常” specifically attracted users interested in or affected by HIV/AIDS, and our findings reflected the discourse patterns of Douyin users engaging with people living with HIV/AIDS daily-life narratives rather than representing the broader Chinese digital public. Additionally, provincial engagement with HIV-related content might not be uniform, as variation in local public health campaigns, disease prevalence, and awareness initiatives might influence comment volume and thematic composition independently of actual public attitudes.

Third, despite using advanced natural language processing techniques, automated text analysis has inherent limitations in capturing linguistic nuances such as sarcasm, irony, and localized dialects. The 6-character filtering threshold may have systematically excluded shorter supportive expressions, such as “加油.” Fourth, reliance on a single platform,

Douyin, introduces data constraints, as its user demographics and algorithmic dynamics may differ from other platforms. IP-based geolocation might not reflect users’ region of primary socialization, particularly for migrant workers, introducing measurement errors in province-level discourse assignment.

Fifth, the SHAP-GIS framework provided a cross-sectional view but did not capture temporal dynamics of public health perceptions. Sixth, the normalization of Topic Index using mobile internet access users might introduce systematic bias if platform-specific user demographics, such as age and urbanicity, vary systematically across provinces. Seventh, while LASSO mitigates overfitting, the small sample size ( $N=31$ ) remained a constraint for the stability of parameter estimates, and our SHAP-GIS localized estimated contributions should be interpreted as exploratory rather than confirmatory.

Several avenues for future inquiry emerge from these limitations. Researchers could integrate digital analysis with offline household surveys to cross-validate findings against more representative population samples. The adoption of multimodal models capable of simultaneously analyzing video content and text would better capture emotional subtexts in short-form video interactions. Multiplatform comparative studies encompassing Weibo, Xiaohongshu, and other social media would offer a more comprehensive view of the national digital health landscape. Finally, incorporating longitudinal data would enable tracking how the “geography of stigma” evolves over time, providing dynamic evidence for place-based public health interventions.

## Conclusions

By integrating LLM-based text classification, LASSO-regularized regression, spatial econometrics, and SHAP-GIS decomposition, this study offered a multidimensional perspective on how HIV/AIDS narratives were consumed and judged across China. This study leveraged LLMs to enable large-scale annotation of social media content related to HIV discourse, enhancing the efficiency of text classification in computational social science research. Methodologically, we integrated “big data” classification with “small sample” provincial analysis, providing a framework for addressing spatial nonstationarity in social media health data. Theoretically, our findings suggested that the associations between sociodemographic and epidemiological factors and digital discourse are geographically contingent rather than universally uniform. Empirically, we identified specific provincial patterns, revealing a complex digital landscape characterized by elevated stigma discourse in southwestern provinces, higher supportive discourse in certain regions, and a north-south demographic association with moral judgment.

## Acknowledgments

The authors sincerely thank the editor and anonymous reviewers for their insightful comments, constructive feedback, and careful evaluation of the manuscript. Their suggestions substantially improved the quality, clarity, and rigor of this study. Generative AI was used solely as an analytical tool for automated text classification of social media comments, as detailed in

the Methods section. Additionally, we used AI-assisted tools to check for grammatical errors and improve readability during the final polishing stage.

### Funding

This work was supported by the National Key Research and Development Program of China (2024YFC3307602) and the Guangdong Provincial Talent Program (2023JC10X009 and 2025D03J0019).

### Data Availability

The raw comment data cannot be publicly shared due to platform terms of service. The cleaned topic-level aggregated data, LASSO results, and prompt templates and core classification code are available on GitHub [35].

### Authors' Contributions

Conceptualization: PZ, YL

Data curation: PZ, YL

Formal analysis: PZ

Funding acquisition: NJ, PH

Investigation: PZ, YL

Methodology: PZ

Project administration: PZ

Resources: PH, NJ

Software: PZ

Supervision: PH, NJ

Validation: ZW, PZ

Visualization: PZ, YL

Writing – original draft: PZ, YL

Writing – review & editing: PZ, YL

### Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Multimedia Appendix 1

Distinguishing stigmatizing from nonstigmatizing content.

[[PDF File \(Adobe File\), 158 KB-Multimedia Appendix 1](#)]

### Multimedia Appendix 2

Confusion matrix for large language model classification validation.

[[PDF File \(Adobe File\), 281 KB-Multimedia Appendix 2](#)]

### Multimedia Appendix 3

Data distribution of candidate variables.

[[PDF File \(Adobe File\), 213 KB-Multimedia Appendix 3](#)]

### Multimedia Appendix 4

Ranking of variable selection frequency in least absolute shrinkage and selection opLeast Absolute Shrinkage and Selection Operator regression.

[[PDF File \(Adobe File\), 194 KB-Multimedia Appendix 4](#)]

### Multimedia Appendix 5

Variance inflation factor test results for the 4 selected variables.

[[PDF File \(Adobe File\), 144 KB-Multimedia Appendix 5](#)]

### Multimedia Appendix 6

Topic Index spatial distribution heat map.

[[PDF File \(Adobe File\), 1083 KB-Multimedia Appendix 6](#)]

### References

1. Payagala S, Pozniak A. The global burden of HIV. Clin Dermatol. 2024;42(2):119-127. [doi: [10.1016/j.clindermatol.2024.02.001](https://doi.org/10.1016/j.clindermatol.2024.02.001)] [Medline: [38387533](https://pubmed.ncbi.nlm.nih.gov/38387533/)]

2. Xu JJ, Han MJ, Jiang YJ, et al. Prevention and control of HIV/AIDS in China: lessons from the past three decades. *Chin Med J (Engl)*. Nov 10, 2021;134(23):2799-2809. [doi: [10.1097/CM9.0000000000001842](https://doi.org/10.1097/CM9.0000000000001842)] [Medline: [34759226](https://pubmed.ncbi.nlm.nih.gov/34759226/)]
3. Zigah EY, Abu-Ba'are GR, Shamrock OW, et al. "For my safety and wellbeing, I always travel to seek health care in a distant facility"—the role of place and stigma in HIV testing decisions among GBMSM – BSGH 002. *Health & Place*. Sep 2023;83:103076. [doi: [10.1016/j.healthplace.2023.103076](https://doi.org/10.1016/j.healthplace.2023.103076)]
4. Laingoen O, Jones CH, Williams S, Apidechkul T, Bishop S. Negotiating risk and resilience: HIV prevention practices among hill tribe communities in northern Thailand. *Health & Place*. Jan 2026;97:103592. [doi: [10.1016/j.healthplace.2025.103592](https://doi.org/10.1016/j.healthplace.2025.103592)]
5. Zeng Q, Yang Y, Zhang L, et al. The impact of the National Syphilis Prevention Program on the prevalence of syphilis among people living with HIV in China: a systematic review and meta-analysis. *J Int AIDS Soc*. Jan 2025;28(1):e26408. [doi: [10.1002/jia2.26408](https://doi.org/10.1002/jia2.26408)] [Medline: [39763073](https://pubmed.ncbi.nlm.nih.gov/39763073/)]
6. Akfırat S, Bayrak F, Üzümcüker E, Ergiyen T, Yurtbakan T, Uysal MS. The roles of social norms and leadership in health communication in the context of COVID-19. *Soc Sci Med*. Apr 2023;323:115868. [doi: [10.1016/j.socscimed.2023.115868](https://doi.org/10.1016/j.socscimed.2023.115868)] [Medline: [36963212](https://pubmed.ncbi.nlm.nih.gov/36963212/)]
7. Zhang P, Wei Z, Kong F. Reconfiguring responsibility: an empirical analysis of crisis discourse and situational crisis communication on Douyin. *Int J Disaster Risk Sci*. Feb 2026;17(1):16-32. [doi: [10.1007/s13753-026-00693-2](https://doi.org/10.1007/s13753-026-00693-2)]
8. Fung A, Hu Y. Douyin, storytelling, and national discourse. *Int Commun Chin Cult*. Dec 2022;9(3-4):139-147. [doi: [10.1007/s40636-022-00259-z](https://doi.org/10.1007/s40636-022-00259-z)]
9. Zhan X, Song M, Shrader CH, Forbes CE, Algarin AB. Decoding HIV discourse on social media: large-scale analysis of 191,972 tweets using machine learning, topic modeling, and temporal analysis. *J Med Internet Res*. Aug 29, 2025;27:e76745. [doi: [10.2196/76745](https://doi.org/10.2196/76745)] [Medline: [40882217](https://pubmed.ncbi.nlm.nih.gov/40882217/)]
10. Gibbs A, Gumede D, Luthuli M, et al. Opportunities for technologically driven dialogical health communication for participatory interventions: perspectives from male peer navigators in rural South Africa. *Soc Sci Med*. Jan 2022;292:114539. [doi: [10.1016/j.socscimed.2021.114539](https://doi.org/10.1016/j.socscimed.2021.114539)]
11. Li YT, Chen ML, Lee HW. Health communication on social media at the early stage of the pandemic: examining health professionals' COVID-19 related tweets. *Social Science & Medicine*. Apr 2024;347:116748. [doi: [10.1016/j.socscimed.2024.116748](https://doi.org/10.1016/j.socscimed.2024.116748)]
12. Zeng L, Zhao Y, Ming WK. Navigating visibility on Weibo among people living with HIV: qualitative study. *J Med Internet Res*. Aug 25, 2025;27:e72490. [doi: [10.2196/72490](https://doi.org/10.2196/72490)] [Medline: [40854227](https://pubmed.ncbi.nlm.nih.gov/40854227/)]
13. Piao Y, Taguchi N, Harada K, et al. Stigma attitudes toward HIV/AIDS from 2011 through 2023 in Japan: retrospective study in Japan. *J Med Internet Res*. May 12, 2025;27:e69696. [doi: [10.2196/69696](https://doi.org/10.2196/69696)] [Medline: [40354646](https://pubmed.ncbi.nlm.nih.gov/40354646/)]
14. Beecroft A, Vaikla O, Engel N, Duchaine T, Liang C, Pant Pai N. Evidence on digital HIV self-testing from accuracy to impact: updated systematic review. *J Med Internet Res*. Mar 4, 2025;27:e63110. [doi: [10.2196/63110](https://doi.org/10.2196/63110)] [Medline: [40053740](https://pubmed.ncbi.nlm.nih.gov/40053740/)]
15. Wray TB, Chan PA, Klausner JD, et al. Using web analytics data to identify platforms and content that best engage high-priority HIV populations in online and social media marketing advertisements. *Digit HEALTH*. 2023;9:20552076231216547. [doi: [10.1177/20552076231216547](https://doi.org/10.1177/20552076231216547)] [Medline: [38025100](https://pubmed.ncbi.nlm.nih.gov/38025100/)]
16. Rzeszutek M, Gruszczyńska E, Pięta M, Malinowska P. HIV/AIDS stigma and psychological well-being after 40 years of HIV/AIDS: a systematic review and meta-analysis. *Eur J Psychotraumatol*. 2021;12(1):1990527. [doi: [10.1080/20008198.2021.1990527](https://doi.org/10.1080/20008198.2021.1990527)] [Medline: [34868481](https://pubmed.ncbi.nlm.nih.gov/34868481/)]
17. Obregon C. A critical assessment of theories/models used in health communication for HIV/AIDS. *J Health Commun*. Jan 2000;5(sup1):5-15. [doi: [10.1080/10810730050019528](https://doi.org/10.1080/10810730050019528)]
18. Han X, Li B, Qu J, Zhu Q. Weibo friends with benefits for people live with HIV/AIDS? The implications of Weibo use for enacted social support, perceived social support and health outcomes. *Social Science & Medicine*. Aug 2018;211:157-163. [doi: [10.1016/j.socscimed.2018.06.016](https://doi.org/10.1016/j.socscimed.2018.06.016)]
19. Operario D, Sun S, Bermudez AN, et al. Integrating HIV and mental health interventions to address a global syndemic among men who have sex with men. *Lancet HIV*. Aug 2022;9(8):e574-e584. [doi: [10.1016/S2352-3018\(22\)00076-5](https://doi.org/10.1016/S2352-3018(22)00076-5)] [Medline: [35750058](https://pubmed.ncbi.nlm.nih.gov/35750058/)]
20. Scarneo SE, Kerr ZY, Kroshus E, et al. The socioecological framework: a multifaceted approach to preventing sport-related deaths in high school sports. *J Athl Train*. Apr 1, 2019;54(4):356-360. [doi: [10.4085/1062-6050-173-18](https://doi.org/10.4085/1062-6050-173-18)]
21. Yin P. Cross-linking geotagged social media data with public health registries for spatial health research. *Health Place*. Mar 2026;98:103621. [doi: [10.1016/j.healthplace.2026.103621](https://doi.org/10.1016/j.healthplace.2026.103621)]
22. Frohlich N, Mustard C. A regional comparison of socioeconomic and health indices in a Canadian province. *Social Science & Medicine*. May 1996;42(9):1273-1281. [doi: [10.1016/0277-9536\(95\)00220-0](https://doi.org/10.1016/0277-9536(95)00220-0)]

23. Valentini I, Nurchis MC, Altamura G, Cicchetti A, Damiani G, Arbia G. The impact of socio-economic conditions on individuals' health: development of an index and examination of its association with three of the most frequently registered diseases in Lazio region of Italy. *Soc Indic Res*. Jul 2024;173(3):691-708. [doi: [10.1007/s11205-024-03354-3](https://doi.org/10.1007/s11205-024-03354-3)]
24. Darin-Mattsson A, Fors S, Kåreholt I. Different indicators of socioeconomic status and their relative importance as determinants of health in old age. *Int J Equity Health*. Sep 26, 2017;16(1):173. [doi: [10.1186/s12939-017-0670-3](https://doi.org/10.1186/s12939-017-0670-3)] [Medline: [28950875](https://pubmed.ncbi.nlm.nih.gov/28950875/)]
25. Pettigrew TF. Intergroup contact theory. *Annu Rev Psychol*. 1998;49:65-85. [doi: [10.1146/annurev.psych.49.1.65](https://doi.org/10.1146/annurev.psych.49.1.65)] [Medline: [15012467](https://pubmed.ncbi.nlm.nih.gov/15012467/)]
26. Chan BT, Tsai AC. Personal contact with HIV-positive persons is associated with reduced HIV-related stigma: cross-sectional analysis of general population surveys from 26 countries in sub-Saharan Africa. *J Int AIDS Soc*. Jan 11, 2017;20(1):21395. [doi: [10.7448/IAS.20.1.21395](https://doi.org/10.7448/IAS.20.1.21395)] [Medline: [28362067](https://pubmed.ncbi.nlm.nih.gov/28362067/)]
27. Sun P, Lu W, Jin L. How the natural environment in downtown neighborhood affects physical activity and sentiment: using social media data and machine learning. *Health Place*. Jan 2023;79:102968. [doi: [10.1016/j.healthplace.2023.102968](https://doi.org/10.1016/j.healthplace.2023.102968)] [Medline: [36628806](https://pubmed.ncbi.nlm.nih.gov/36628806/)]
28. Wang S, Liang C, Gao Y, et al. Social media insights into spatio-temporal emotional responses to COVID-19 crisis. *Health Place*. Jan 2024;85:103174. [doi: [10.1016/j.healthplace.2024.103174](https://doi.org/10.1016/j.healthplace.2024.103174)]
29. Wei Z, Xie Y, Xiao D, Zhang S, Hui P, Zhou M. Social media discourses on interracial intimacy: tracking racism and sexism through Chinese geo-located social media data. Presented at: Proceedings of the ACM Web Conference 2024. Association for Computing Machinery. 2337-2346; Singapore. May 13, 2024. URL: <https://dl.acm.org/doi/proceedings/10.1145/3589334> [Accessed 2026-02-26] [doi: [10.1145/3589334.3645334](https://doi.org/10.1145/3589334.3645334)]
30. Zhang P, Zhang H, Kong F. Research on online public opinion in the investigation of the “7–20” extraordinary rainstorm and flooding disaster in Zhengzhou, China. *International Journal of Disaster Risk Reduction*. Apr 2024;105:104422. [doi: [10.1016/j.ijdrr.2024.104422](https://doi.org/10.1016/j.ijdrr.2024.104422)]
31. Chen J, Geng Y, Chen Z, et al. Zero-shot and few-shot learning with knowledge graphs: a comprehensive survey. *Proc IEEE*. 2023;111(6):653-685. [doi: [10.1109/JPROC.2023.3279374](https://doi.org/10.1109/JPROC.2023.3279374)]
32. Song Y, Wang T, Cai P, Mondal SK, Sahoo JP. A comprehensive survey of few-shot learning: evolution, applications, challenges, and opportunities. *ACM Comput Surv*. Dec 31, 2023;55(13s):1-40. [doi: [10.1145/3582688](https://doi.org/10.1145/3582688)]
33. Guo D, Yang D, Zhang H, et al. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature*. Sep 2025;645(8081):633-638. [doi: [10.1038/s41586-025-09422-z](https://doi.org/10.1038/s41586-025-09422-z)] [Medline: [40962978](https://pubmed.ncbi.nlm.nih.gov/40962978/)]
34. Xu Q, Liu Y, Wang D, Huang S. Automatic recognition of cross-language classic entities based on large language models. *npj Herit Sci*. 2025;13(1). [doi: [10.1038/s40494-025-01624-y](https://doi.org/10.1038/s40494-025-01624-y)]
35. GitHub. URL: <https://github.com/cauzp/JMIR93007> [Accessed 2026-09-03]
36. DeJonckheere M, Vaughn LM, James TG, Schondelmeyer AC. Qualitative thematic analysis in a mixed methods study: guidelines and considerations for integration. *J Mix Methods Res*. Jul 2024;18(3):258-269. [doi: [10.1177/15586898241257546](https://doi.org/10.1177/15586898241257546)]
37. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012;2012:276-282. [doi: [10.11613/BM.2012.031](https://doi.org/10.11613/BM.2012.031)]
38. Zhang P, Wei Z, Zhou M, Evans J, Hui P. Beyond aggregate feedback: a spatially-aware computational framework for understanding citizen-state interaction on social media. *Trans Soc Comput*. Mar 31, 2026;9(1):1-22. [doi: [10.1145/3790251](https://doi.org/10.1145/3790251)]
39. Muthukrishnan R, Rohini R. LASSO: a feature selection technique in predictive modeling for machine learning. Presented at: 2016 IEEE International Conference on Advances in Computer Applications (ICACA); Oct 24, 2016:18-20; Coimbatore, India. [doi: [10.1109/ICACA.2016.7887916](https://doi.org/10.1109/ICACA.2016.7887916)]
40. O'Brien RM. A caution regarding rules of thumb for variance inflation factors. *Qual Quant*. Sep 11, 2007;41(5):673-690. [doi: [10.1007/s11135-006-9018-6](https://doi.org/10.1007/s11135-006-9018-6)]
41. Zhang C, Luo L, Xu W, Ledwith V. Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. *Science of The Total Environment*. Jul 2008;398(1-3):212-221. [doi: [10.1016/j.scitotenv.2008.03.011](https://doi.org/10.1016/j.scitotenv.2008.03.011)]
42. Uyanık GK, Güler N. A study on multiple linear regression analysis. *Procedia - Social and Behavioral Sciences*. Dec 2013;106:234-240. [doi: [10.1016/j.sbspro.2013.12.027](https://doi.org/10.1016/j.sbspro.2013.12.027)]
43. Hayes AF, Cai L. Using heteroskedasticity-consistent standard error estimators in OLS regression: an introduction and software implementation. *Behav Res Methods*. Nov 2007;39(4):709-722. [doi: [10.3758/bf03192961](https://doi.org/10.3758/bf03192961)] [Medline: [18183883](https://pubmed.ncbi.nlm.nih.gov/18183883/)]
44. Rüttenauer T. Spatial regression models: a systematic comparison of different model specifications using Monte Carlo Experiments. *Sociological Methods & Research*. May 2022;51(2):728-759. [doi: [10.1177/0049124119882467](https://doi.org/10.1177/0049124119882467)]

45. Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*. Mar 2002;27(1):77-83. [doi: [10.3102/10769986027001077](https://doi.org/10.3102/10769986027001077)]
46. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Presented at: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4-9, 2017:4768-4777; Long Beach, California, USA. URL: <https://doi.org/10.48550/arXiv.1705.07874> [Accessed 2026-02-24]
47. Zhang J, Ma X, Zhang J, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. *J Environ Manage*. Apr 2023;332:117357. [doi: [10.1016/j.jenvman.2023.117357](https://doi.org/10.1016/j.jenvman.2023.117357)]
48. Zhang P, Li Y, Wei Z, Hui P. The geography of climate concern: a large-scale analysis of public discourse on extreme heat in China using social media and explainable AI. *Environ Impact Assess Rev*. Mar 2026;117:108227. [doi: [10.1016/j.eiar.2025.108227](https://doi.org/10.1016/j.eiar.2025.108227)]
49. Zhang P, Jiang N, Li Y, Wei Z, Hui P. Leveraging large language models to explain provincial heterogeneity in online discussion of water pollution in China: a socioeconomic and environmental perspective. *J Clean Prod*. Feb 2026;543:147517. [doi: [10.1016/j.jclepro.2026.147517](https://doi.org/10.1016/j.jclepro.2026.147517)]
50. Vaughan E, Tinker T. Effective health risk communication about pandemic influenza for vulnerable populations. *Am J Public Health*. Oct 2009;99 Suppl 2(Suppl 2):S324-32. [doi: [10.2105/AJPH.2009.162537](https://doi.org/10.2105/AJPH.2009.162537)] [Medline: [19797744](https://pubmed.ncbi.nlm.nih.gov/19797744/)]
51. Tian J, Zhang R. Moral judgments influence emotional responses and comment lengths through the moderating role of linguistic style matching. *Sci Rep*. 2025;15(1). [doi: [10.1038/s41598-025-07079-2](https://doi.org/10.1038/s41598-025-07079-2)]
52. Stackpool-Moore L, Logie CH, Cloete A, Reygan F. What will it take to get to the heart of stigma in the context of HIV? *J Int AIDS Soc*. Jul 2022;25 Suppl 1(Suppl 1):e25934. URL: <https://onlinelibrary.wiley.com/toc/17582652/25/S1> [doi: [10.1002/jia2.25934](https://doi.org/10.1002/jia2.25934)] [Medline: [35818934](https://pubmed.ncbi.nlm.nih.gov/35818934/)]
53. Tian W, Ge J. Decoding the apple paradox: a critical discourse analysis of gender, technology, and nationalism in China's digital space. *Humanit Soc Sci Commun*. 2024;11(1). [doi: [10.1057/s41599-024-03819-3](https://doi.org/10.1057/s41599-024-03819-3)]
54. Price SF, Butler S, Mocarski R. "The world wants us dead": stigma and the social construction of health in *Pose*. *Critical Studies in Media Communication*. Aug 8, 2021;38(4):307-320. [doi: [10.1080/15295036.2021.1934503](https://doi.org/10.1080/15295036.2021.1934503)]
55. Li L, Wu Z, Liang LJ, et al. Reducing HIV-related stigma in health care settings: a randomized controlled trial in China. *Am J Public Health*. Feb 2013;103(2):286-292. [doi: [10.2105/AJPH.2012.300854](https://doi.org/10.2105/AJPH.2012.300854)] [Medline: [23237175](https://pubmed.ncbi.nlm.nih.gov/23237175/)]
56. Alonzo AA, Reynolds NR. Stigma, HIV and AIDS: an exploration and elaboration of a stigma trajectory. *Soc Sci Med*. Aug 1995;41(3):303-315. [doi: [10.1016/0277-9536\(94\)00384-6](https://doi.org/10.1016/0277-9536(94)00384-6)]
57. Iveniuk J, Calzavara L, Bullock S, et al. Social capital and HIV-serodiscordance: disparities in access to personal and professional resources for HIV-positive and HIV-negative partners. *SSM Popul Health*. Mar 2022;17:101056. [doi: [10.1016/j.ssmph.2022.101056](https://doi.org/10.1016/j.ssmph.2022.101056)] [Medline: [35342785](https://pubmed.ncbi.nlm.nih.gov/35342785/)]
58. Bingley WJ, Greenaway KH, Haslam SA. A Social-Identity Theory of Information-Access Regulation (SITAR): understanding the psychology of sharing and withholding. *Perspect Psychol Sci*. May 2022;17(3):827-840. [doi: [10.1177/1745691621997144](https://doi.org/10.1177/1745691621997144)] [Medline: [34606731](https://pubmed.ncbi.nlm.nih.gov/34606731/)]
59. Rutledge JD. Exploring the role of empowerment in Black women's HIV and AIDS activism in the United States: an integrative literature review. *Am J Community Psychol*. Jun 2023;71(3-4):491-506. [doi: [10.1002/ajcp.12644](https://doi.org/10.1002/ajcp.12644)] [Medline: [36609979](https://pubmed.ncbi.nlm.nih.gov/36609979/)]

## Abbreviations

**FDR:** false discovery rate  
**GDP:** gross domestic product  
**GIS:** geographic information system  
**LASSO:** least absolute shrinkage and selection operator  
**LLM:** large language model  
**MLR:** multiple linear regression  
**OLS:** ordinary least squares  
**RQ:** research question  
**SDM:** spatial Durbin model  
**SEM:** spatial error model  
**SHAP:** Shapley Additive Explanations  
**SLM:** spatial lag model  
**UNAIDS:** Joint United Nations Programme on HIV/AIDS

*Edited by Ivan Steenstra; peer-reviewed by Albert Whata, Defu Yuan, Dillon Chrimes, Zhaohui Su; submitted 06.Feb.2026; final revised version received 01.Sep.2026; accepted 01.Sep.2026; published 09.Oct.2026*

Please cite as:

Zhang P, Li Y, Wei Z, Jiang N, Hui P

*Spatial Heterogeneity of Digital HIV/AIDS Stigma in China: Cross-Sectional Ecological Study*

*J Med Internet Res* 2026;28:e93007

URL: <https://www.jmir.org/2026/1/e93007>

doi: [10.2196/93007](https://doi.org/10.2196/93007)

© Pu Zhang, Yiliang Li, Zheng Wei, Na Jiang, Pan Hui. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 09.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.