Scientific Research Publishing

Open Journal of Applied Sciences, 2026, 16(11), 4683-4702
https://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

Quantitative Evaluation of Cyber Literacy among Active Weibo Users in Shanghai: An Entropy-Weight TOPSIS Approach

Rui Yang

School of Management, University of Shanghai for Science and Technology, Shanghai, China
Email: 15955320214@163.com

How to cite this paper: Yang, R. (2026) Quantitative Evaluation of Cyber Literacy among Active Weibo Users in Shanghai: An Entropy-Weight TOPSIS Approach. *Open Journal of Applied Sciences*, **16**, 4683-4702.
https://doi.org/10.4236/ojapps.2026.1611258

Received: October 23, 2026
Accepted: November 21, 2026
Published: November 24, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

With the rapid development of information technology, cyber literacy has become an increasingly prominent research focus as a measure of individuals' ability to engage with information in digital environments. Traditional questionnaire-based assessments are limited by subjectivity and susceptibility to social desirability bias. This study therefore develops a quantitative assessment model using social media behavioral data to provide an exploratory assessment of cyber literacy among 598 active ordinary Weibo users in Shanghai. Drawing on the New Media Literacy Scale framework, the assessment comprises 12 indicators across four dimensions: functional consumption, critical consumption, functional production, and critical production. Indicator weights are determined using the entropy-weight method, comprehensive literacy scores are calculated using the technique for order preference by similarity to ideal solution, and feature-weighted K-means clustering is used to classify users into literacy levels. The results identify the proportion of external information citations, the proportion of rich-media use, and the proportion of rumor-debunking and information-verification behaviors as key indicators differentiating users' cyber literacy. Clustering yields high, moderate, and low literacy groups. The low-literacy users exhibit obvious characteristics such as reliance on a single information source and a lack of critical thinking. By shifting from cognitive self-assessment to behavioral measurement, this study extends traditional questionnaire-based assessment approaches and provides a quantitative basis for the exploratory assessment of cyber literacy in a specific population.

Keywords

Cyber Literacy, NMLS Framework, Risk Assessment, Entropy-Weight TOPSIS, K-Means Clustering

1. Introduction

As national background, the 57th CNNIC report states that by December 2025 China had 1.125 billion Internet users and an Internet penetration rate of 80.1% [1]. As social media continues to expand, problems such as the proliferation of misinformation [2] [3], frequent online harassment, and information cocoons reinforced by algorithmic recommendations have become increasingly salient. These developments pose substantial challenges to internet users' ability to evaluate information, think critically, and uphold ethical standards online. As a key competence for citizens' participation and development in the digital age, the assessment and enhancement of cyber literacy have become important issues in cyberspace governance. A cross-sectional survey of university students in Malaysia reported that technical and cognitive dimensions of digital literacy were positively associated with mitigating misinformation, whereas socio-emotional literacy showed an opposite association [4].

The concept of network literacy was introduced by McClure in 1994 [5], who distinguished between knowledge and skills and defined its scope primarily in terms of individuals' ability to use networked tools to obtain and process information. Questionnaire surveys have been used in academic assessments of media literacy competences [6]. Although this approach has produced substantial advances in theoretical development and scale construction, it faces several limitations in practice. First, self-report questionnaires are susceptible to social desirability bias: respondents tend to overestimate their literacy levels, creating discrepancies between assessment results and actual abilities. Second, questionnaire surveys generally have difficulty covering large samples and provide one-time, static cross-sectional data, making it difficult to capture users' actual behavioral characteristics in dynamic online environments. Finally, traditional literacy assessments tend to emphasize cognition while overlooking behavioral performance in practical operations and interactions, resulting in a gap between knowledge and action [6].

Against this background, this study aims to address the limitations of traditional assessment paradigms by developing a quantitative model for assessing users' cyber literacy using social media behavioral data. It adopts the New Media Literacy Scale (NMLS) framework proposed by Koc and Barut [7], designs a quantitative indicator system tailored to the characteristics of Weibo, applies entropy-weight TOPSIS to conduct a comprehensive multidimensional assessment of users' literacy, and uses K-means clustering to classify literacy levels and identify risks. The study extends the methodological repertoire for cyber literacy assessment and provides empirical evidence to inform the governance of online platform content ecosystems and targeted literacy education [8].

2. Literature Review

2.1. Research on Cyber Literacy

In recent years, with the rapid development of information technology, cyber lit-

eracy has become an increasingly prominent research focus as a measure of individuals' ability to engage with information in digital environments. Scholars in China and internationally have extensively discussed its definition and conceptual scope. The concept of network literacy was introduced by McClure, who distinguished between knowledge and skills and defined its scope primarily in terms of individuals' ability to use networked tools to obtain and process information [5]. A systematic review describes social media literacy as extending media literacy to digital platforms and identifies critical thinking, socio-emotional, and technical competences as central components [9]. A single-dimensional measure cannot fully characterize adolescents' media literacy; multidimensional, comprehensive analysis building on competence assessment is therefore necessary [6]. These studies provide theoretical support for a multidimensional conceptualization of cyber literacy and a basis for subsequent assessment and intervention.

2.2. Evolution of Cyber Literacy Assessment Methods

Current approaches to cyber literacy assessment can be broadly classified into three categories: subjective self-assessment using questionnaire scales, objective measurement through controlled experiments, and quantitative assessment through behavioral data analysis. Questionnaire surveys remain the predominant approach, with a relatively mature assessment paradigm. Researchers have developed instruments including internet literacy scales for high-school and junior students [10] [11] and the New Media Literacy Scale [7], advancing systematic, multidimensional assessment of cyber literacy. Young (2015) examined a self-reported media-literacy assessment instrument in the context of social work students and educators [12]. However, questionnaire-based self-assessment is affected by social desirability bias, as respondents often overestimate their literacy levels, producing systematic discrepancies between assessment results and actual abilities [6]. Moreover, such scales primarily measure cognitive self-perceptions rather than actual behavioral performance, revealing a pronounced gap between knowledge and action [13]. Studies have also examined links between news media literacy, critical-thinking dispositions, and the evaluation of misinformation [14] [15]. For experimental approaches, settings differ fundamentally from real-world online environments, limiting ecological validity. High design costs also constrain large-scale implementation, while data collected at a single time point cannot dynamically capture changes in literacy levels [16]. As big data technologies have matured, some scholars have begun advocating behavioral data-based assessment, representing an important paradigm shift from cognitive self-assessment to behavioral measurement [17] [18]. Digital traces left by users during everyday information interactions include records of reposting, commenting, original content creation, and information citation. These traces can provide objective indications of literacy in practice and offer notable advantages in data volume, timeliness, and objectivity. Using publicly available Facebook user data for factor analysis, Ahn identified four behavioral dimensions and specific correspondences be-

tween behaviors and literacy [19]. However, research on the transfer from traditional questionnaire-based assessment to social media-based assessment remains limited.

2.3. Risk Assessment and Early Warning

With the widespread dissemination of fake news and misinformation, systems for cyber literacy risk assessment and early warning have become increasingly important in an era of information overload. Previous studies have proposed methods for detecting fake news on social media, highlighted the risks associated with social networks during information dissemination, and developed pattern-driven fake news detection strategies based on network analysis, thereby broadening perspectives on risk assessment [2] [3]. Shu *et al.* also examined explicit and implicit social-media user-profile features for fake-news detection, reporting their potential usefulness for distinguishing fake from real news [20]. Social network analysis and public opinion monitoring can complement literacy-oriented risk assessment, as illustrated by research on emergency rumor management [21]. In comprehensive multi-indicator assessment and risk identification, entropy-weight TOPSIS has been widely applied because of its methodological rigor and objectivity. The entropy-weight method determines indicator weights using information entropy, avoiding the subjective arbitrariness of manually assigned weights. TOPSIS, or the technique for order preference by similarity to ideal solution, ranks assessed entities by calculating their distances from positive and negative ideal solutions and is suitable for comprehensive assessment of multidimensional data [22]. For example, Jiang Changbin and colleagues applied entropy-weight TOPSIS to assess online public opinion risks in higher education institutions [8]. In addition, K-means clustering is an unsupervised learning method that automatically groups samples according to their underlying data characteristics and is commonly used for user segmentation and risk-level classification. Combining entropy-weight TOPSIS with K-means clustering can support both precise quantitative scoring and systematic tiered early warning, offering a suitable technical approach to cyber literacy assessment.

3. Design of the Assessment Indicator System

Given the characteristics of publicly available social media data, the first-level indicators in this study are theoretically grounded in the New Media Literacy Scale framework. This framework organizes individuals' literacy practices in new media environments into four dimensions: functional consumption, critical consumption, functional production, and critical production. It provides a systematic account of users' overall abilities in online information acquisition, analysis and judgment, content expression, and critical thinking. Building on this framework, the study operationalizes these abstract literacy dimensions in relation to Weibo's behavioral sequence of following, browsing, reposting, commenting, and publishing, constructing a system of 12 second-level indicators. These indicators were

selected primarily according to the principles of theoretical correspondence, behavioral observability, data quantifiability, and platform compatibility. The different dimensions of cyber literacy are thus mapped onto objective variables extractable from users' social media behaviors, enabling behavior-based measurement of cyber literacy. Details are presented in Figure 1.

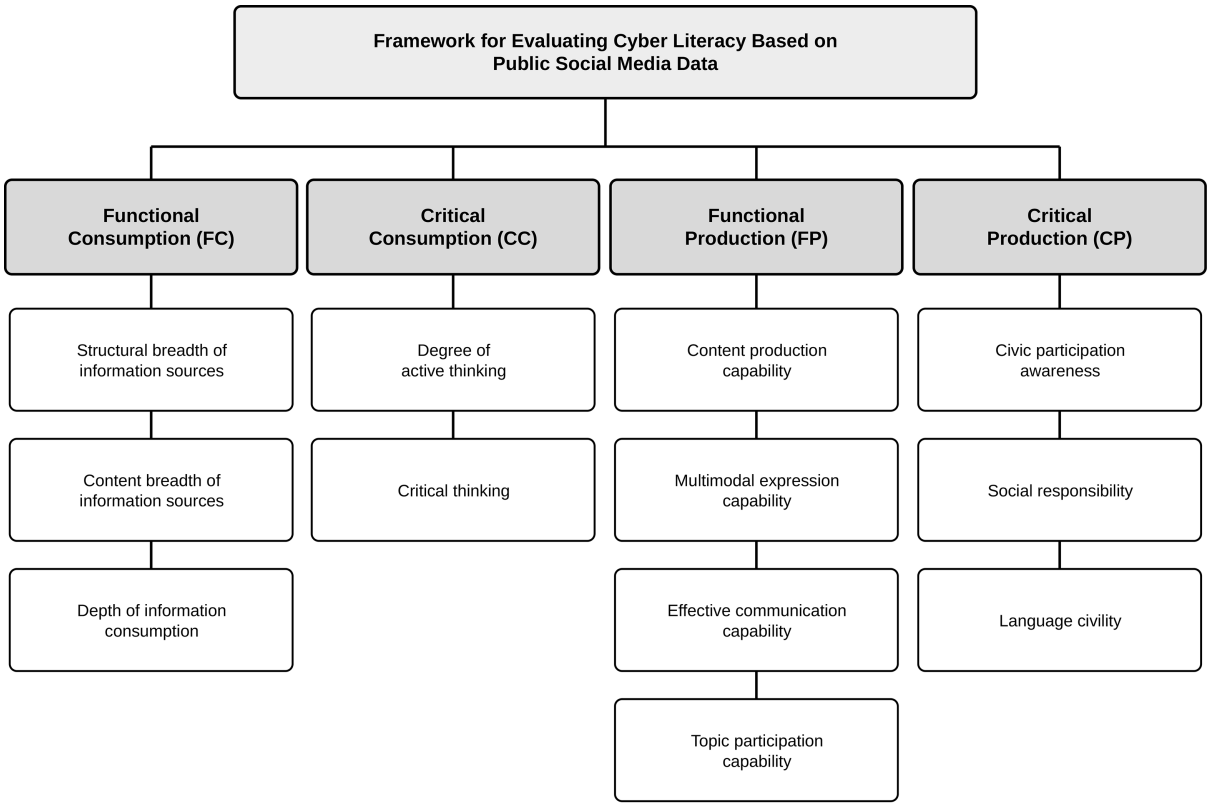


Figure 1. Framework of the cyber literacy assessment indicator system based on publicly available social media data.

3.1. Functional Consumption

The functional consumption dimension primarily assesses users' technical abilities to acquire, understand, and integrate information in digital environments, together with the breadth of their access to resources, and constitutes a foundational dimension of cyber literacy assessment. It reflects individuals' initiative in moving beyond information cocoons in an environment of information overload. This study quantifies this dimension at three levels: the structural breadth of information sources, content breadth, and depth of information consumption.

Structural breadth is quantified by analyzing the entropy of the distribution of different social roles in users' following lists, including official media, self-media accounts, key opinion leaders, and ordinary users. The structural distribution of information sources shapes the diversity of perspectives to which users are exposed. High entropy values are interpreted as indicating an ability to obtain information across social strata and organizational boundaries, making this measure important for assessing whether users maintain a balanced information ecosystem

and avoid the pitfalls of reliance on a single type of source.

For content breadth, this study uses the diversity of topics followed by users to reflect their ability to construct interdisciplinary knowledge maps in digital spaces. A latent Dirichlet allocation topic model is used to semantically cluster the profile descriptions and tags of followed accounts, and the number of topics covered is calculated. Broader content coverage suggests that users are better able to use social media for multidimensional observation of society and have greater potential for information integration.

In addition, the depth of information consumption reflects users' habit of tracing sources and seeking in-depth justifications. According to the holistic digital literacy model proposed by Eshet, individuals should possess the ability to cross-check information from multiple sources [23]. In this study, we measure users' capacity to escape the platform's closed algorithmic environment and verify information on other platforms by calculating the proportion of external references such as URLs and academic literature within their Weibo posts.

3.2. Critical Consumption

The critical consumption dimension assesses users' abilities to analyze, question and evaluate incoming information, aiming to measure individuals' capacity to judge the authenticity of social media content. This study quantifies this dimension using two key indicators: the degree of active thinking and critical thinking.

For active thinking, quantification is based on the proportion of reposts accompanied by user comments. Within social media diffusion mechanisms, one-click reposting often reflects blind conformity or emotional impulses. In contrast, adding a comment usually implies that the user has conducted cognitive processing and viewpoint review before sharing. This ratio indirectly reflects users' awareness of deliberation during information dissemination and serves as an important indicator for evaluating users' critical thinking.

3.3. Functional Production

Functional production reflects individuals' ability to transition from information consumers to prosumers and assesses users' technical proficiency in creating, processing, and disseminating content using new media tools. This study comprehensively evaluates this ability through four aspects: content production, multimodal expression, effective dissemination, and topic participation.

For the indicator of content production capacity, the proportion of original Weibo posts is adopted as the measurement basis. A high share of original posts indicates that users are more willing to compose texts independently and express their own viewpoints. Multimodal expression capacity examines users' ability to convey information through symbolic carriers such as images and videos. Simply uploading materials does not equal competence in this dimension. The core lies in whether users can flexibly allocate various media resources according to their communication purposes, which constitutes an essential part of multimodal liter-

acy [24]. Many users turn to rich media content either to strengthen the appeal of their statements or to convey complicated information hard to describe in plain text. Such choices reflect their understanding of media attributes [25]. Online communication today is no longer limited to text-only forms and has moved toward multimodal integration. It places higher requirements on users: they need more than basic skills to edit pictures or short videos. They should select proper expressive forms matched to the message they intend to deliver. Such competence represents users' comprehensive ability to operate media proficiently and improve the efficiency of information transmission. Accordingly, this study uses the proportion of rich media posts to characterize users' multimodal expression capacity.

For effective dissemination ability, the dissemination performance of content serves as an objective criterion for assessing production quality. This study calculates an effective dissemination score by weighting the numbers of likes, comments, and reposts received by users' posts, using this score to reflect their understanding of audience psychology and their application of community rules. Topic participation ability additionally reflects users' engagement in the digital public sphere [26] and is quantified through the frequency of hashtag use. Frequent and accurate use of hashtags suggests users' technical awareness of how to leverage platform mechanisms to participate in public issues and join specific interest communities, reflecting their ability to collaborate and integrate within digital communities.

3.4. Critical Production

The critical production dimension assesses users' sense of social responsibility, ethical standards, and civic awareness during content production and sharing, representing the highest behavioral expression of cyber literacy. This study conducts a quantitative assessment at three levels: civic participation awareness, social responsibility, and civility of language.

At the level of civic participation, the assessment examines users' attention to public affairs and willingness to participate, reflecting their sense of digital citizenship responsibility and critical reflection on social issues [27]. The study uses text classification to identify the proportion of posts addressing public issues, thereby quantifying users' civic participation awareness through the "proportion of posts addressing social issues"; detailed calculations are provided in the appendix. This indicator reflects the extent to which users connect personal expression with social development and serves as an important criterion for assessing whether individuals can use online platforms to exercise their rights to social participation and demonstrate digital citizenship awareness.

In the dimension of social responsibility, we adopt the proportion of rumor refutation and information verification behaviors as the indicator to count how frequently users post rumor-correcting content and label questionable information. Fowler-Watt and McDougall present media literacy as a means of building resilience to disinformation and frame verification as a responsibility extend-

ing beyond journalists [28]. Users who take the initiative to clarify false information and exercise caution before sharing unconfirmed content demonstrate a willingness to sustain a healthy online ecosystem. Such foresight regarding communication consequences and self-restraint not only signals a sense of social responsibility, but also marks an important way critical thinking translates into constructive action.

Civility of language is the only negative indicator within this evaluation framework. Drawing on a sentiment dictionary, this study calculates the share of Weibo posts containing hostile words such as insults and personal attacks among all sampled posts; detailed calculation procedures are provided in the appendix. This indicator gauges users' boundaries of speech and moral bottom line. It reflects individual media literacy as well as recognition and compliance with norms of online civility, and serves as a reference for identifying online misconduct and assessing the harmony of cyberspace.

4. Construction of the Assessment Model

4.1. Entropy-Based Indicator Weighting

The entropy-weight method is an objective weighting method that measures the dispersion of indicators by calculating their information entropy. Greater dispersion corresponds to lower information entropy and a larger amount of information provided, and therefore warrants a greater weight.

Let n denote the number of users and m the number of assessment indicators. The original assessment matrix is $X = \left(x_{ij}\right)_{n\times m}$, where x_{ij} is the value of indicator j for user i ($i = 1,\cdots,n; j = 1,\cdots,m$). To remove differences in units and value ranges, the indicators are normalized to obtain $Y = \left(y_{ij}\right)_{n\times m}$. Benefit-type and cost-type indicators are normalized using Equations (1) and (2), respectively.

$$y_{ij} = \frac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}} \tag{1}$$

$$y_{ij} = \frac{\max_i x_{ij} - x_{ij}}{\max_i x_{ij} - \min_i x_{ij}} \tag{2}$$

The information entropy and objective weight of each indicator are calculated using Equations (3) and (4).

$$e_j = -\frac{1}{\ln(n)}\sum_{i=1}^{n} p_{ij}\ln\left(p_{ij}\right) \tag{3}$$

$$w_j = -\frac{d_j}{\sum_{j=1}^{m} d_j} \tag{4}$$

The proportions, divergence coefficients, and weight constraints are defined as follows:

$$p_{ij} = -\frac{y_{ij}}{\sum_{i=1}^{n} y_{ij}}, d_j = 1 - e_j, 0 \le w_j \le 1, \sum_{j=1}^{m} w_j = 1.$$

4.2. Calculation of Composite Scores Using TOPSIS

The TOPSIS method evaluates the relative performance of alternatives by constructing positive and negative ideal solutions for a multi-attribute problem and calculating each alternative's distances from these ideal solutions.

First, the assessment matrix is normalized to obtain $X' = \left(x'_{ij}\right)_{n\times m}$. The normalized matrix is then weighted using the entropy-based indicator weights to obtain the weighted matrix in Equation (5).

$$R = \left(r_{ij}\right)_{n\times m}, r_{ij} = w_j \cdot x'_{ij} \tag{5}$$

The positive and negative ideal solutions are determined from the maximum and minimum weighted values for each indicator across the n users, as shown in Equations (6) and (7).

$$S_j^+ = \max_{1\le i\le n} r_{ij} \tag{6}$$

$$S_j^- = \min_{1\le i\le n} r_{ij} \tag{7}$$

Each user's Euclidean distances from the positive and negative ideal solutions are calculated using Equations (8) and (9), respectively.

$$Sep_j^+ = \sqrt{\sum_{j=1}^{m}\left(r_{ij} - S_j^+\right)^2} \tag{8}$$

$$Sep_j^- = \sqrt{\sum_{j=1}^{m}\left(r_{ij} - S_j^-\right)^2} \tag{9}$$

Finally, Equation (10) calculates the relative closeness coefficient. A higher value indicates that the user is closer to the positive ideal solution and farther from the negative ideal solution, corresponding to higher cyber literacy.

$$C_i = \frac{Sep_i^-}{Sep_i^+ + Sep_i}, C_i \in [0,1]. \tag{10}$$

5. Empirical Analysis

5.1. Data Sources and Indicator Quantification

Data collection and sample screening in this study followed the principles of public accessibility, representativeness, and control of heterogeneity, and adopted a purposive design combining topic-based recruitment with snowball referral, with the aim of constructing a sample that reflects objectively the behavioral characteristics of cyber literacy within a defined population. To exclude systematic errors in language habits and in literacy assessment arising from differences in regional culture and living environment, the sample was restricted to Shanghai, residence being determined jointly from participation in Shanghai-topic communities and from the IP location displayed by the platform for the sampled posts. The study further focused on active ordinary social media users, selecting those with between 100 and 500 followers and defining activity as sustained public content production: registration for at least twelve months, at least one original post or repost in every month of the observation window, and at least thirty publicly visible posts

in total. In the sampling procedure, the researchers first browsed the Weibo "Shanghai Super Topic" in real time to identify users actively participating in discussions of local issues, and then observed interaction ties to select other local users associated with these participants through a snowball approach. A total of 681 candidates meeting these preliminary criteria were identified. To ensure completeness across data dimensions and accuracy of assessment, users for whom complete behavioral data could not be obtained were excluded, as were non-individual accounts whose profile descriptions or verification information indicated organizational operation, leaving 598 Weibo users as the final sample. The data include (1) user metadata (2) relationship-network data (3) content-generation data and (4) content-interaction data. All data were obtained through publicly accessible platform interfaces; to protect participants' privacy and reduce the risk of personal identification, data processing strictly followed the Personal Information Protection Law of the People's Republic of China and related personal information protection regulations, and identity-related information was de-identified: usernames, user IDs, and profile links were deleted or replaced, sensitive information unnecessary for the research was removed, and only the behavioral variables required for the quantitative assessment of cyber literacy were retained. Illustrative records from the dataset are presented in Table 1.

Table 1. Illustrative records from the original Weibo dataset.

User no.	Post date	Post text (English translation)	Reposts	Comments	Likes	Post type
1	2020-04-10	#HaidilaoApologizes# I haven't been there in a long time… The first priority in running a business is to make money, not to do charity. Can we all agree on that?	0	1	0	Original post
2	2021-04-05	Paying tribute from afar [candle]	152	0	0	Repost
3	2023-03-03	#eStar [Super Topic] Have they lost their composure?	0	36	98	Original post
4	2022-03-29	Fuck, I'm laughing so hard I could throw up. p3 Lao Yu is already deep in meditation and hasn't joined in at all.jpg	1	0	0	Repost
5	2022-12-29	NetEase's videos are pretty well made, very sentimental… Chinese media are really boring; few of them have any brains.	1	0	0	Original post
6	2022-06-30	Repost	0	0	0	Repost
…	…	…	…	…	…	…

Note. The full analytical sample comprises 598 users.

5.2. Entropy-Based Indicator Weights

After preprocessing the publicly available social media data of the 598 Weibo users, the data were standardized, and the entropy-weight method was used to cal-

culate the weights of the 12 cyber literacy assessment indicators, as shown in Table 2. Further details are provided in the supplementary materials.

Table 2. Entropy-based weights of cyber literacy indicators.

Indicator	Information entropy (e_j)	Divergence coefficient (d_j)	Objective weight (w_j)
Proportion of external information citations	0.6263	0.3737	0.2867
Proportion of rich-media use	0.6318	0.3682	0.2825
Proportion of rumor-debunking and information-verification behaviors	0.8491	0.1509	0.1158
Content quality score	0.8924	0.1076	0.0826
Frequency of hashtag use	0.8956	0.1044	0.0801
Number of topic categories represented by followed accounts	0.9312	0.0688	0.0528
Proportion of posts addressing social issues	0.9506	0.0494	0.0379
Frequency of critical vocabulary	0.9525	0.0475	0.0364
Proportion of original Weibo posts	0.9859	0.0141	0.0109
Proportion of reposts with added comments	0.9878	0.0122	0.0094
Distribution of followed account types	0.9959	0.0041	0.0032
Proportion of posts containing aggressive language	0.9976	0.0024	0.0018

Note. N = 598. For indicator j, e_j denotes information entropy, $d_j = 1 - e_j$ is the divergence coefficient, and w_j is the entropy-based objective weight. Indicators are listed in descending order of weight. The distribution of followed account types is quantified using entropy. Aggressive language is the only cost-type indicator. Values are retained at the source precision; the rounded weights sum to 1.0001.

This study adopts the entropy-weight method as an objective weighting approach, using dispersion to determine the magnitude of a weight: the greater an indicator's variation among users, the lower its information entropy, the more information it provides for distinguishing among users, and thus the greater its weight. In this study, the three indicators with the highest weights are the proportion of external information citations (0.2867), the proportion of rich-media use (0.2825), and the proportion of rumor-debunking and information-verification behaviors (0.1158); these are the behaviors on which the 598 users in this sample differ most, and which therefore contribute most to distinguishing among users in the composite score and in the clustering, and they are also the behaviors that mark the boundary between active users in Shanghai. The low weights of the proportion of original posts (0.0109) and the proportion of aggressive language (0.0018) reflect that the two are near-universal in this sample, and their discriminating capacity is therefore weak, but this does not diminish their conceptual relevance.

5.3. Cyber Literacy Scores

Using the entropy-weight TOPSIS method, the standardized data were entered

into the formulas to calculate each user's relative closeness to the ideal solution for cyber literacy, namely, the cyber literacy score. The resulting distribution of cyber literacy scores is shown in Figure 2. Descriptive statistics are summarized in Table 3.

Table 3. Descriptive statistics of TOPSIS-based cyber literacy scores.

Statistic	TOPSIS score (C_i)	Interpretation
Minimum	0.1637	Lowest cyber literacy score in the sample
Maximum	0.7197	Highest cyber literacy score in the sample
Mean	0.4466	Average cyber literacy score in the sample
Median	0.4450	Middle value of the sample score distribution
First quartile (Q_1)	0.4067	Score below which approximately 25% of users fall
Third quartile (Q_3)	0.4845	Score below which approximately 75% of users fall
Standard deviation (SD)	0.0784	Dispersion of cyber literacy scores in the sample

Note. N = 598. C_i is the TOPSIS relative closeness coefficient, ranging from 0 to 1; higher values indicate higher cyber literacy.

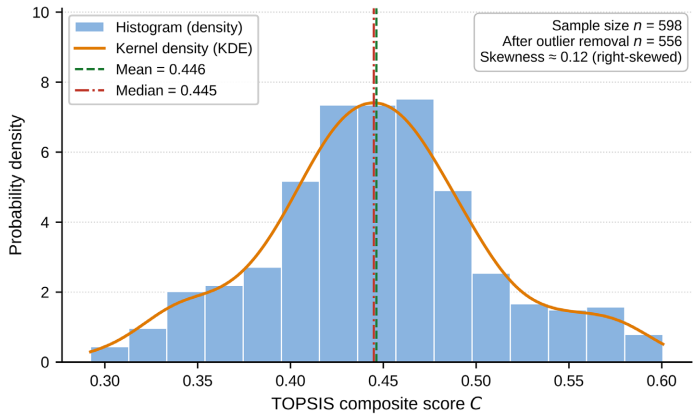


Figure 2. Distribution of TOPSIS-C cyber literacy scores.

5.4. Clustering Results and Risk Assessment

To ensure that the cluster analysis adequately reflects the contribution of the core competency dimensions, the present study adopted the feature-weighted K-means clustering method. The composite score C derived from the TOPSIS calculation was used as the clustering input; C ranged from 0.1637 to 0.7197, with a mean of 0.4466, and the number of clusters K was determined jointly by the elbow method and the silhouette coefficient [29]. With K increasing from 2 to 8, K-means clustering was performed on C, yielding within-cluster sum of squared errors (SSE) of 1.7498, 0.8467, 0.5075, 0.3583, 0.2445, 0.1743 and 0.1371, respectively, with reductions between adjacent values of K of 51.61%, 40.06%, 29.40%, 31.75%, 28.73% and 21.36%. The reduction was greatest when K increased from 2 to 3 and narrowed rapidly thereafter, producing a distinct elbow in the curve at K = 3 (Figure 3).

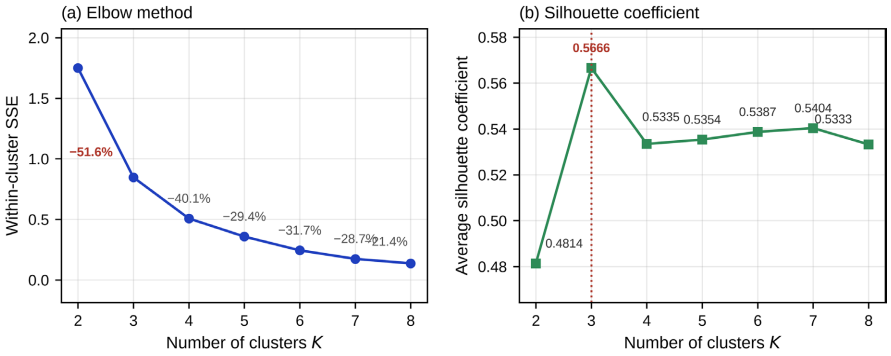


Figure 3. Determination of the number of clusters. (a) Elbow method applied to the composite TOPSIS score. (b) Average silhouette coefficient.

Three cluster centers corresponding to cyber literacy levels were ultimately obtained, dividing users into three cyber literacy categories: high literacy, moderate literacy, and low literacy. The score intervals for these three levels were determined using the distances between users' TOPSIS scores and the cluster centers, as shown in Table 4.

Table 4. Classification of users by cyber literacy level.

Cyber literacy level	Cluster center	TOPSIS score interval	Users (n)	Percentage (%)
High	0.5747	[0.5120, 0.7197]	89	14.88
Moderate	0.4493	[0.3933, 0.5120]	395	66.05
Low	0.3373	[0.1637, 0.3933]	114	19.07

Note. N = 598.

The distribution of users across the three cyber literacy levels is shown in Figure 4.

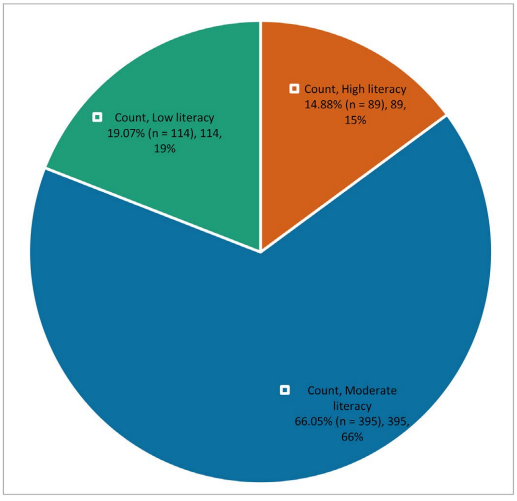


Figure 4. Distribution of cyber literacy levels among active Weibo users in Shanghai based on TOPSIS.

Based on the clustering results above and the cyber literacy indicator data, this study calculated the mean value of each indicator for the high-, moderate-, and low-literacy groups. A radar chart illustrating differences in cyber literacy indicators across the three levels was then plotted, as shown in Figure 5.

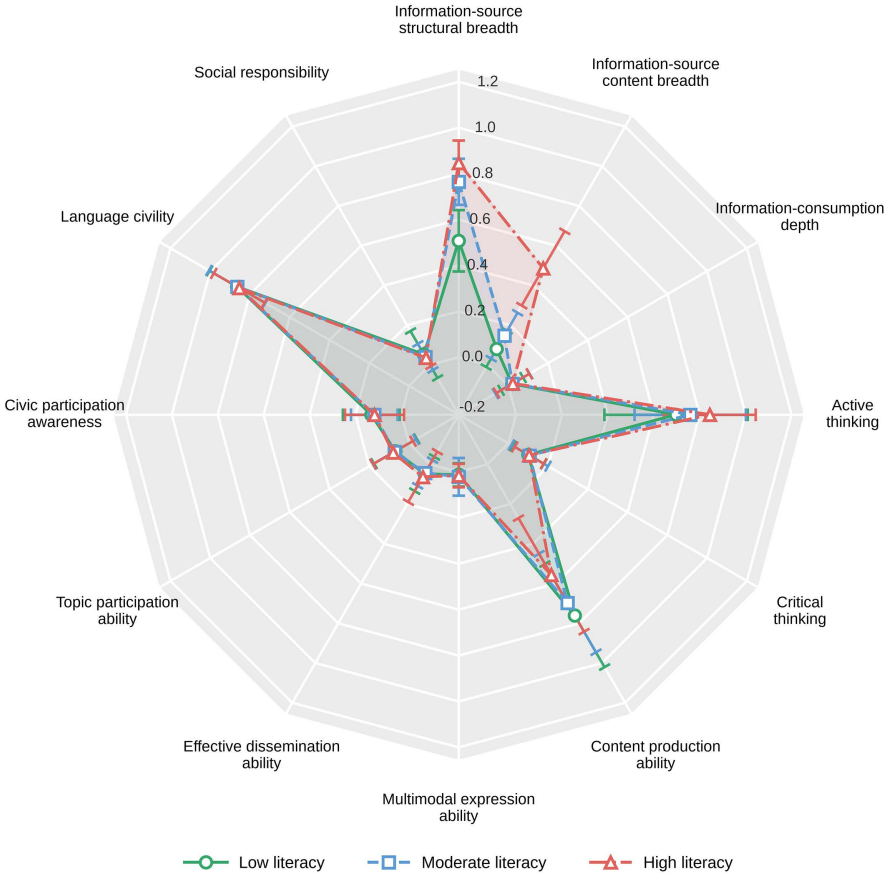


Figure 5. Cyber literacy indicator profiles of the high-, moderate-, and low-literacy groups across the 12 assessment indicators.

5.4.1. High-Literacy Users

The high-literacy group comprised 89 users, accounting for 14.88% of the total sample. Its cluster center for the TOPSIS composite score was 0.5747, placing it among the minority of users with high literacy.

This group shows the most prominent tendency to cite external information, indicating that they are more likely to actively step out of the information-enclosed environment of a single platform and break through the filter bubble. In the dimension of critical consumption, they also demonstrate stronger rational analysis and questioning capabilities, which enables them to resist emotional herd communication effectively.

Notably, this group performs worse in content production capacity than users in the other two tiers. This reflects their characteristic of prudent communication. Such behavioral features can be interpreted through theories such as impression management and evaluation anxiety. Goffman's dramaturgical theory suggests

that individuals engage in self-presentation in social interactions to maintain an ideal image [30]. Users with high literacy are more inclined to regard social media as a "front stage" for maintaining their image and reputation. In such a highly transparent environment, they tend to speak cautiously to avoid image crises [31]. Meanwhile, evaluation anxiety makes high-literacy users set higher standards for the quality and influence of their own remarks, and they are more sensitive to others' evaluations, which leads them to remain silent in public spaces [32]. In addition, people with high literacy scores possess critical thinking and a sense of responsible communication, which implicitly raises the threshold for content creation; as a result, they rarely gain substantial interactions on social platforms. Therefore, high-literacy users do not take the absolute lead across all indicators. Instead, they effectively combine critical consumption competence with responsible communication behaviors, allowing them to defend against information risks and generate positive externalities for cyberspace.

5.4.2. Moderate-Literacy Users

The medium-literacy group has 395 people, or 66.05% of the sample. Their clustering center on the TOPSIS composite score is 0.4493, making them the largest group in the sample. The core trait of this group is "strong on function, weak on critical thinking," which reflects an imbalance in their ability structure. They are skilled at using platforms and willing to create content, and they also have a reasonably wide range of information sources.

However, their high activity and productivity are not matched by equally strong critical thinking. They score lower than the high-literacy group on active thinking, and there is a gap between how much they share and how carefully they check before posting.

So this is a group that can spread information easily but whose judgment is relatively unstable. Under algorithm-driven feeds, their high activity could amplify the spread of misinformation. Platforms should therefore use in-feed reminders, fact-check entry points, and targeted guidance to help them better tell true from false and share more carefully.

5.4.3. Low-Literacy Users

The low-literacy group has 114 people, or 19.07% of the sample. Their clustering center on the TOPSIS composite score is 0.3373, the lowest of the three groups, which makes them the group that needs the most attention in identifying network literacy risks.

Their behavior shows a single information source, weak critical thinking, and reposting without checking first, forming a passive "little processing, fast spreading" pattern. Notably, this group is strongly motivated to create content, even more than the medium- and high-literacy users. As a result, they struggle to tell rumors and false information from facts. In an environment where algorithm recommendations and emotional sharing reinforce each other, they are more easily misled and can unknowingly become spreaders of false information.

For this group, governance and intervention should focus more on protection and correction than on pure policy measures. On one hand, platforms can lower the chance of them spreading misinformation through anti-rumor warnings, prioritizing authoritative sources in feeds, and downgrading content suspected to be false. On the other hand, this group should be a key target of digital literacy education and platform rule guidance, to gradually build up their basic ability to judge information and share it carefully.

6. Conclusions

This study constructs a quantitative evaluation model of network literacy based on the entropy-weight TOPSIS method and conducts an empirical analysis of Weibo users. The results show that the proportion of external information citation, the proportion of rich media use, and the proportion of fact-checking and information-verification behaviors carry relatively large weights in the network literacy score, and are identified as key indicators affecting literacy levels. The overall literacy distribution of the sample is slightly skewed, with highly literate users accounting for only 14.88%. This research realizes a paradigm shift from "cognitive self-assessment" to "behavioral measurement", and provides an operable approach for the objective measurement of network literacy using publicly available behavioral data from social platforms.

Li and Kuang [33] note that big-data- and AI-enabled public-opinion management can provide new governance resources and tools, while also raising risks such as algorithmic misrepresentation, misinformation, information cocoons, privacy concerns, and data abuse. At the platform governance level, the governing approach should shift from post hoc content governance to proactive risk alerts. Platforms could integrate the literacy assessment model developed in this study into their content recommendation mechanisms, replacing uniform rules with user literacy profiles to enable tiered, targeted interventions. Incorporating core indicators with high weights, such as the proportion of external information citations and the proportion of rumor-debunking and information-verification behaviors, into credibility scores for user-generated content could move risk identification upstream. Positive incentives could also encourage high-literacy users to participate voluntarily in maintaining the platform ecosystem, creating a governance arrangement in which platform intervention and community self-governance complement one another.

At the literacy education level, educational resources should be directed more strongly toward the critical dimensions. The moderate-literacy group, which accounted for 66% of the sample, generally exhibited a structural imbalance characterized by proficient functional operation but weak critical judgment. This suggests that the foremost gap in current literacy education is not the reinforcement of technical skills, but the systematic cultivation of critical reading and information-verification capabilities. In two open online consultations, young people's perspectives suggested that digital literacy can support information verification,

while effective responses to misinformation also need to address platform design and wider structural constraints [34].

This study has several limitations that point to directions for future research. First, multidimensional sample expansion is needed. The sample was restricted to active Weibo users in Shanghai. Although this offers some advantages in controlling for regional cultural differences, the highly specific sample limits the nationwide generalizability of the findings. Future research should expand geographical coverage to include cities at different tiers and rural areas, and extend the analysis to platforms with diverse content ecosystems, such as Douyin and Xiaohongshu, to examine the robustness and generalizability of the findings. Second, deep semantic models could be introduced. The current critical consumption dimension relies on shallow word-frequency statistics. Pretrained language models such as BERT and RoBERTa could be used for in-depth semantic analysis and sentiment polarity identification in user-generated content, enabling semantic-level refinement of indicators in the critical dimensions. Future studies could examine whether hybrid semantic models, including the BERT-LSTM approach explored for misinformation detection and digital-literacy education [35], improve semantic-level analysis of user-generated content. Third, dynamic literacy tracking is needed. This study provides a static, cross-sectional assessment. Future research could incorporate time-series analysis or longitudinal panel data models to track changes in users' literacy levels over time and explore the effects of major public events on literacy-related behaviors and the mechanisms of subsequent recovery, thereby providing a more timely scientific basis for real-time platform governance. Fourth, validation using multiple methods should be introduced. Because the available data were anonymous and retrospective, this study could not conduct questionnaire-based follow-up with the same participants. Future research could combine behavioral tracking and questionnaire-based self-assessment for a subset of participants to evaluate the model's incremental validity in addressing the knowledge-behavior gap. Fifth, the indicator weights in this study are derived from the entropy weight method, and subjective weighting methods such as the analytic hierarchy process are not adopted. For future research, a combined weighting scheme integrating the entropy weight method and expert weighting can be adopted, so that both information volume and normative judgment can jointly participate in weight determination.

During manuscript preparation, the author used AI-assisted tools for translation, language polishing, and revision of selected passages. The core research was conducted and the study data were obtained independently by the author.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] China Internet Network Information Center (CNNIC) (2026) The 57th Statistical Re-

port on China's Internet Development. China Internet Network Information Center. (In Chinese)
https://www.cnnic.net.cn/NMedia-File/2026/0304/MAIN1772588317069TUXN3827X8.pdf

[2] Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H. (2017) Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, **19**, 22-36. https://doi.org/10.1145/3137597.3137600

[3] Zhou, X. and Zafarani, R. (2019) Network-Based Fake News Detection: A Pattern-Driven Approach. *ACM SIGKDD Explorations Newsletter*, **21**, 48-60. https://doi.org/10.1145/3373464.3373473

[4] Anthonysamy, L. and Sivakumar, P. (2024) A New Digital Literacy Framework to Mitigate Misinformation in Social Media Infodemic. *Global Knowledge*, *Memory and Communication*, **73**, 809-827. https://doi.org/10.1108/gkmc-06-2022-0142

[5] McClure, C.R. (1994) Network Literacy: A Role for Libraries? *Information Technology and Libraries*, **13**, 115-125. https://eric.ed.gov/?id=EJ486787

[6] Pereira, S. and Moura, P. (2022) Assessing Media Literacy Competences: Reflections and Recommendations from a Quantitative Study. *Journal of Media Literacy Education*, **14**, 79-93. https://doi.org/10.23860/jmle-2022-14-3-7

[7] Koc, M. and Barut, E. (2016) Development and Validation of New Media Literacy Scale (NMLS) for University Students. *Computers in Human Behavior*, **63**, 834-843. https://doi.org/10.1016/j.chb.2016.06.035

[8] Jiang, C., Xu, Z., Wang, H. and Xu, S. (2024) Risk Assessment and Early Warning of Online Public Opinion on University Teacher Ethics Based on the Entropy Weight TOPSIS Method. *Information Science*, **42**, 113-120. (In Chinese)
http://som.whut.edu.cn/rsgz/szdw/201307/t20130726_1292016.shtml

[9] Polanco-Levicán, K. and Salvo-Garrido, S. (2022) Understanding Social Media Literacy: A Systematic Review of the Concept and Its Competences. *International Journal of Environmental Research and Public Health*, **19**, Article 8807. https://doi.org/10.3390/ijerph19148807

[10] Ma, S., Wang, Y., Shu, Z., Duan, Z. and Sun, L. (2024) Development and Validation of Internet Literacy Scale for High School Students. *Education and Information Technologies*, **29**, 1427-1454. https://doi.org/10.1007/s10639-023-11641-8

[11] Huang, Y., Liu, H., Wang, W., Dong, R. and Tang, Y. (2021) The Junior Students' Internet Literacy Scale: Measure Development and Validation. *International Journal of Environmental Research and Public Health*, **18**, Article 10120. https://doi.org/10.3390/ijerph181910120

[12] Young, J.A. (2015) Assessing New Media Literacies in Social Work Education: The Development and Validation of a Comprehensive Assessment Instrument. *Journal of Technology in Human Services*, **33**, 72-86. https://doi.org/10.1080/15228835.2014.998577

[13] Potter, W.J. (2010) The State of Media Literacy. *Journal of Broadcasting & Electronic Media*, **54**, 675-696. https://doi.org/10.1080/08838151.2011.521462

[14] Ku, K.Y.L., Kong, Q., Song, Y., Deng, L., Kang, Y. and Hu, A. (2019) What Predicts Adolescents' Critical Thinking about Real-Life News? the Roles of Social Media News Consumption and News Media Literacy. *Thinking Skills and Creativity*, **33**, Article ID: 100570. https://doi.org/10.1016/j.tsc.2019.05.004

[15] Orhan, A. (2023) Fake News Detection on Social Media: The Predictive Role of University Students' Critical Thinking Dispositions and New Media Literacy. *Smart*

Learning Environments, **10**, Article No. 29.
https://doi.org/10.1186/s40561-023-00248-8

[16] Valle, N., Zhao, P., Freed, D., Gorton, K., Chapman, A.B., Shea, A.L., *et al.* (2025) Towards a Critical Framework of Social Media Literacy: A Systematic Literature Review. *Review of Educational Research*, **95**, 701-746.
https://doi.org/10.3102/00346543241247224

[17] Papaioannou, T. (2011) Assessing Digital Media Literacy among Youth through Their Use of Social Networking Sites. *Revista de Informatică Socială*, **8**, 37-48.
http://www.ris.uvt.ro/wp-content/uploads/2011/09/tao.pdf

[18] Literat, I. (2014) Measuring New Media Literacies: Towards the Development of a Comprehensive Assessment Tool. *Journal of Media Literacy Education*, **6**, 15-27.
https://doi.org/10.23860/jmle-6-1-2

[19] Ahn, J. (2013) What Can We Learn from Facebook Activity? Using Social Learning Analytics to Observe New Media Literacy Skills. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, Leuven, 8-13 April 2013, 135-144. https://doi.org/10.1145/2460296.2460323

[20] Shu, K., Zhou, X., Wang, S., Zafarani, R. and Liu, H. (2019) The Role of User Profiles for Fake News Detection. *Proceedings of the* 2019 *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Vancouver, 27-30 August 2019, 436-439. https://doi.org/10.1145/3341161.3342927

[21] Wang, X., Lin, J. and Wang, X. (2025) Public Opinion Network Structure and Rumor Management in Emergencies Based on Social Media User Data. *Journal of Systems & Management*, **34**, 1342-1354. (In Chinese)
https://xtglxb.sjtu.edu.cn/EN/abstract/abstract1951.shtml

[22] Yang, D.Q. and Han, Q.Y. (2020) Fuzzy Multi-Index Comprehensive Evaluation Model of Online Public Opinion Situation Grades Based on TOPSIS Method. *Journal of Systems Science and Mathematical Sciences*, **40**, 1352-1364. (In Chinese)
https://doi.org/10.12341/jssms13928

[23] Eshet-Alkalai, Y. (2012) Thinking in the Digital Era: A Revised Model for Digital Literacy. *Issues in Informing Science and Information Technology*, **9**, 267-276.
https://doi.org/10.28945/1621

[24] Kress, G. (2010) Multimodality: A Social Semiotic Approach to Contemporary Communication. Routledge. https://doi.org/10.4324/9780203970034

[25] van Leeuwen, T. (2005) Introducing Social Semiotics. Routledge.

[26] Jenkins, H. (2006) Convergence Culture: Where Old and New Media Collide. New York University Press.

[27] Kellner, D. and Share, J. (2007) Critical Media Literacy, Democracy, and the Reconstruction of Education. In: Macedo, D. and Steinberg, S.R., Eds., *Media Literacy: A Reader*, Peter Lang, 3-23.
https://cxarchive.gseis.ucla.edu/xchange/critical-uses-of-media-and-technology/xpress/critical-media-literacy-democracy-and-the-reconstruction-of-education

[28] Fowler-Watt, K. and McDougall, J. (2019) Media Literacy versus Fake News: Fact Checking and Verification in the Era of Fake News and Post-Truths. *Journalism Education: The Journal of the Association for Journalism Education*, **8**, 59-68.
https://eprints.bournemouth.ac.uk/32851/

[29] Milligan, G.W. and Cooper, M.C. (1985) An Examination of Procedures for Determining the Number of Clusters in a Data Set. *Psychometrika*, **50**, 159-179.
https://doi.org/10.1007/bf02294245

[30] Goffman, E. (1959) The Presentation of Self in Everyday Life. Doubleday.

[31] Ryan, F.V.C., Cruickshank, P., Hall, H. and Lawson, A. (2020) Blurred Reputations: Managing Professional and Private Information Online. *Journal of Librarianship and Information Science*, **52**, 16-26. https://doi.org/10.1177/0961000618769977

[32] Gurgun, S., Arden-Close, E., Phalp, K. and Ali, R. (2023) Online Silence: Why Do People Not Challenge Others When Posting Misinformation? *Internet Research*, **33**, 1928-1948. https://doi.org/10.1108/intr-06-2022-0407

[33] Li, M.D. and Kuang, Y. (2021) Network Public Opinion Management in the Context of Big Data and Artificial Intelligence: Value, Risk and Path Exploration. *Journal of Beijing University of Technology* (*Social Sciences Edition*), **21**, 1-10. (In Chinese) https://doi.org/10.12120/bjutskxb202106001

[34] Diepeveen, S. and Pinet, M. (2022) User Perspectives on Digital Literacy as a Response to Misinformation. *Development Policy Review*, **40**, e12671. https://doi.org/10.1111/dpr.12671

[35] Wang, J., Wang, X. and Yu, A. (2025) Tackling Misinformation in Mobile Social Networks a BERT-LSTM Approach for Enhancing Digital Literacy. *Scientific Reports*, **15**, Article No. 1118. https://doi.org/10.1038/s41598-025-85308-4