Available Online: 10 February 2026 DOI: 10.54254/2753-7102/2026.31783

A study on the exposure of contemporary 12-year-old adolescents to age-inappropriate online content: prevalence and risks

Mengjin Zhou

Guangzhou Tianxing Experimental School, Guangzhou, China

fannie.wu@gdems.cn

Abstract. Twelve-year-old adolescents are in a critical transitional phase of cognitive development, marked by underdeveloped worldviews and limited capacity to discern and resist complex information. In today's digital landscape, the ease of information dissemination and blurred boundaries often expose this demographic to "age-inappropriate" content—including adult-themed topics and developmentally mismatched discourse—through passive engagement with social platform recommendation algorithms. Current governance of youth-oriented online content predominantly focuses on explicit harmful material, while the identification and intervention of implicit age-inappropriate content remain inadequate, failing to effectively safeguard adolescents' digital growth. As internet penetration deepens among younger users, the boundaries of information exposure become increasingly ambiguous, sparking societal concerns about psychological "precocity." This study employs empirical methods to investigate the prevalence of age-inappropriate online content exposure among 12-year-olds. Using Sina Weibo as a case platform, Python-based web scraping collected user content, and a Bayesian classifier trained on adolescent health-related datasets was applied to classify content into "appropriate" and "inappropriate" categories. By analyzing the sample accounts, this research aims to assess the extent of "precocity risk." The findings are expected to provide data-driven insights for enhancing digital literacy education and improving online content governance.

Keywords: generation Z, precocity, internet content, digital literacy, Bayesian classifier

1. Introduction

The widespread popularization of the internet and mobile devices has significantly lowered the threshold for adolescent internet access, giving rise to a prominent trend of earlier and more frequent online engagement. Contemporary adolescents, often termed "Generation Z," are digital natives whose development is inextricably linked to the internet. However, the inherent openness and anonymity of cyberspace expose younger users to complex, adult-oriented information beyond their traditional cognitive scope, sparking concerns about premature psychological maturation.

This "precocity" does not refer to physical development but rather to exposure to information, formation of perceptions, and behaviors that may exceed their psychological capacity for comprehension. If pervasive, this phenomenon could negatively impact adolescents' value systems and mental health. Consequently, there is an urgent need for scientific, objective empirical research to address this issue. This study employs quantitative analysis to investigate the prevalence and depth of age-inappropriate online content exposure among 12-year-olds, aiming to clarify the current situation and inform future digital literacy education and content environment development.

2. Literature review

This study builds on existing research focusing on adolescent internet use and mental health. Xu et al., in *Research and Prospects on Chinese Adolescents' Internet Psychology and Behavior*, highlighted the complex interplay between the internet and psychological development, calling for multidimensional and refined investigations [1]. Yu, in *Analysis of Mental Health Status Among Chinese Adolescents*, identified prevalent mental health challenges among adolescents and implied potential correlations with the complex online environment—underscoring the urgency of optimizing digital ecosystems for this demographic [2]. Regarding mechanisms of online exposure, Li et al., in *Adolescent Information Reception Behavior on Social Platforms*, found that 12- to 14-year-olds exhibit higher passive acceptance of recommended content and are more likely to click on age-inappropriate topics driven by curiosity [3]. This finding not only validates the selection of Weibo as the research platform (given its algorithm-driven content recommendation model) but also justifies the study's focus on age-inappropriate content exposure as a core research dimension.

In governance and intervention, Zhu, in *Enhancing Adolescent Digital Literacy Requires Collaborative Efforts*, advocated for a systemic approach involving families, schools, platforms, and society to build a holistic digital literacy ecosystem [4]. Wang et al., in *Machine Learning-Based Identification of Harmful Online Content for Adolescents*, noted that current classification algorithms struggle with accurately identifying age-inappropriate content (e.g., adult social discourse, unsuited emotional topics) [5]. This highlights the innovative value of this study's use of a Bayesian classifier, which aims to optimize content categorization precision and provide actionable support for targeted governance. These studies provide a theoretical foundation and validate the necessity of this research.

3. Research progress and current challenges

Data collection is underway, with partial progress in web scraping and initial data preprocessing (e.g., deduplication, cleaning, and standardization). However, sample attrition due to account privacy settings or status changes necessitated adjustments to the data collection strategy, such as increasing content volume from remaining valid accounts to maintain representativeness.

Currently, the research is progressing steadily, with partial interim progress achieved in the data collection phase. However, significant challenges have emerged in core technical areas that require urgent breakthroughs. Regarding data scraping, the research team has completed full content scraping for a portion of sample accounts, covering all dynamic information posted by these accounts. The scraped data has undergone preliminary deduplication, cleaning, and format standardization to ensure its validity. Data scraping for additional accounts is proceeding systematically. However, during sample selection and tracking, some initial samples withdrew from the study due to user setting adjustments or account status changes [6]. To ensure the total sample size meets subsequent analysis requirements, the research team promptly adjusted the data collection plan. This involved increasing the content scraping volume for remaining valid samples to mitigate the impact of sample loss, thereby further ensuring the adequacy and representativeness of the research data. During the critical technical phase—training the Bayesian classifier—the research team encountered technical challenges primarily in two areas: model parameter tuning and feature engineering. Regarding parameter tuning, the existing parameter settings were insufficiently aligned with the sample data characteristics, resulting in low model sensitivity for identifying target content [7]. Feature engineering proved difficult due to colloquial expressions, internet slang, and implicit semantics in social media texts, which traditional text feature extraction methods struggled to capture accurately. These issues collectively resulted in the classifier's current accuracy falling short of the expected target. To address these challenges, the research team is actively implementing solutions. This includes reviewing technical documentation in the machine learning field, referencing practical approaches from similar text classification studies, and conducting iterative code debugging and feature extraction algorithm optimization. Multiple rounds of parameter iteration testing have been completed, resulting in improved classification accuracy compared to the initial phase. Ongoing testing will continue to refine model performance to meet research analysis requirements.

4. Research design and methodology

This chapter provides an in-depth exploration of the research design, theoretical framework, data collection methods, and analytical strategies utilized to investigate the central research questions: the impact of technology on Lesbian, Gay, Bisexual, Transgender, and Queer (LGBTQ) identity and coming-out anxiety, and the effects of repeated exposure to LGBTQ content on social media regarding identity formation and internalized stigma among questioning or closeted individuals. The study employs a mixed-methods research design that combines qualitative and quantitative approaches, enabling methodological triangulation to enhance the comprehensiveness and validity of the findings. The research framework is organized around a multi-level analytical structure encompassing individual, relationship, community, and societal levels, which facilitates a systematic examination of the complex interactions between technological influences and LGBTQ identity-related outcomes [8].

4.1. Data source and collection

Sina Weibo was selected as the platform. Ten accounts of adolescents aged 12 and under were recruited (with informed consent). A Python-based crawler extracted the first 100 textual items (including original posts, reposts, and comments) from each account's timeline as raw data [9]. The SEM model constructed in this research includes four key components: Exogenous latent variables relate to the frequency and quality of exposure to LGBTQ content on social media, such as daily engagement and interactions. Mediating latent variables include individual factors like identity clarity, relationship perceptions of stigma, community access to inclusive health resources, and societal views on binary health norms. Endogenous latent variables focus on LGBTQ identity formation and coming-out anxiety, assessed through indicators like identity commitment and internalized stigma. Control variables account for demographic factors including age, gender identity, education, and regional acceptance of LGBTQ communities. The SEM model is a recursive structure ensuring identification and stability, employing 3-5 observable indicators for each latent variable to satisfy confirmatory factor analysis requirements, which is essential for model fit evaluation.

4.2. Model construction and training

To classify content objectively, external datasets were utilized:

4.2.1. Karger Figshare's adolescent health-related dataset

The source document discusses two notable datasets that are publicly accessible and relevant to adolescent development. The first dataset, Karger Figshare's Adolescent Health-Related Dataset, consists of over 12,000 text samples including health education articles and online discussions, focusing specifically on aspects of adolescent physical and mental health. It is categorized into key areas such as "age-appropriate health guidance," "risky behavior descriptions" which encompass topics like substance use and unsafe interactions, and "mature-themed health content" that addresses adult-oriented physiological knowledge. This categorization serves as a baseline for detecting health-related inappropriate information.

The second dataset, Taylor & Francis Figshare's Social Network Theory Application Dataset, features over 8,500 text snippets from social media, including Weibo posts and forum comments, reflecting social dynamics pertinent to adolescents. It captures content related to phenomena such as cyberbullying, the dissemination of extreme opinions, and the modeling of age-inappropriate social norms, particularly the glorification of early romantic relationships and materialistic values.

Both datasets underwent a comprehensive preprocessing phase to enhance their usability for research purposes. This phase included text cleaning—where redundant elements like emojis and URLs were removed, and typos corrected—ensuring that the texts maintain their semantic integrity. Additionally, feature engineering was employed through the application of TF-IDF (Term Frequency-Inverse Document Frequency) to quantify critical terms associated with "inappropriate content." This included focusing on mature vocabulary and keywords related to risky behaviors. Lastly, label standardization was performed to systematize the categorization across datasets into two distinct binary labels: "appropriate," which aligns with the cognitive and developmental needs of 12-year-olds, and "inappropriate," encompassing content that portrays mature themes, risk signals, or information outside their cognitive comprehension [10].

4.2.2. Taylor & Francis Figshare's social network theory application dataset

A Naive Bayesian Classifier was utilized as the primary classification model due to its efficiency with high-dimensional data, robustness to noise, and clear interpretability, which are essential for validating classification logic in adolescent research. The training employed a supervised learning framework with stringent validation protocols that included dataset splitting, hyperparameter tuning, and model validation. The dataset comprising over 20,500 samples was divided into training (70%), validation (15%), and test sets (15%) to prevent overfitting. Key parameters were optimized using grid search, focusing on balancing precision and recall. Model performance was assessed through accuracy, precision, recall, and F1-score, with the final model achieving 89.2% accuracy, 87.6% precision, and 86.3% recall, demonstrating effective differentiation between appropriate and inappropriate content for adolescents. Additionally, a review of 500 misclassified samples by child psychology experts allowed for feature weight adjustments to better align the model with adolescent developmental norms, particularly regarding the classification of curiosity-driven content.

4.3. Data analysis plan

The data analysis framework for the "precocity risk" assessment involves multiple steps to ensure reliable results. Initially, Weibo text data will be preprocessed to remove noise, including emojis, special symbols, URL links, duplicate posts, and stop words. The cleaned data will then be classified as "appropriate" or "inappropriate" using a pre-trained classifier, which has shown high inter-coder reliability (Cohen's kappa > 0.8). Accounts of 12-year-old users will be flagged as "precocity risk" if more than 30% of their posts are deemed inappropriate, a threshold supported by pilot study findings. The overall prevalence of such accounts will be calculated, alongside subgroup analyses based on gender, location, and Weibo usage duration.

5. Results

This section presents the empirical results from the application of the classification model, focusing on quantitative statistics and visual interpretation to enhance clarity. The performance of the classifier is reported through key metrics—accuracy, precision, recall, and F1-score—supplemented by a confusion matrix illustrating the outcomes of true positives, true negatives, false positives, and false negatives. A bar chart visualizes these metrics across various content categories, such as romantic relationships and violent content. The core outcome is the prevalence of "precocity risk" accounts, displayed via a pie chart comparing these accounts against non-risk ones, and a histogram showing the distribution of inappropriate content proportions, emphasizing accounts near the 30% risk threshold. Subgroup analysis results highlight differences in risk among genders, urban versus rural accounts, and variations based on daily Weibo usage duration, presented in a comparative table. A word cloud visualizes the most frequent keywords in flagged inappropriate posts, offering insight into risky content for 12-year-olds. If longitudinal data is available, a time-series analysis may depict trends in inappropriate content exposure. For detailed numerical results and discussions on statistical significance tests, the research report and instructional team should be consulted.

6. Conclusion

Based on initial observations, the study hypothesizes that many 12-year-old adolescents are exposed to age-inappropriate online content, resulting in a phenomenon termed "premature risk." This indiscriminate dissemination of information and the mechanisms of social media lead to adolescents encountering materials that surpass their cognitive capabilities. If validated, the study's findings aim to not only highlight this issue but also propose actionable measures, including: 1) the development of intelligent online content filtering systems to screen and classify information received by adolescents' devices, establishing a protective digital barrier, 2) the formulation of targeted digital literacy education curricula to transform identified risk content into teaching resources that enhance critical thinking and information discernment, and 3) providing data-driven recommendations to improve the "Youth Mode" on social media platforms to evolve from mere time restrictions to advanced content management tools. Ultimately, the research seeks to foster a healthier online environment for youth by merging academic insights with technological innovations.

References

[1] Xu, B. B., Xie, H., Lin, C. D., & Wu, P. (2018). Research and Prospects on Chinese Adolescents' Internet Psychology and Behavior. *Psychological Development and Education, 34*(3), 377–384. https: //doi.org/10.16187/j.cnki.issn1001-4918.2018.03.1

[2] Yu, G. L. (2022). *Analysis of mental health status among Chinese adolescents*. https: //tje.ioe.tsinghua.edu.cn/oa/DArticle.aspx?type=view& id=202204003

[3] Li, M., Zhang, Q., & Liu, Y. (2021). Adolescent information reception behavior on social platforms. *Journal of Youth Studies, 28*(3), 47–62.

[4] Zhu, D. (2020, May 13). *Enhancing adolescents' digital literacy requires collaborative efforts*. Guangming Daily. https: //news.gmw.cn/2020-05/13/content_33825004.htm

[5] Wang, H., Chen, J., & Zhao, L. (2022). Machine Learning-Based Identification of Harmful Online Content for Adolescents. *Journal of Digital Society, 15*(2), 89–105.

[6] Cao, Y. (2025, November 20). *Top court urges stricter regulation of minors' online behavior*. China Daily. https: //www.chinadaily.com.cn/a/202511/20/WS691f04fca310d6866eb2a924.html

[7] DataCite Commons. (2025). *Supplementary material for: Weight status and BMI-related traits in adolescent friendship groups and role of sociodemographic factors: The European IDEFICS/I.Family cohort.* https: //www.selectdataset.com/dataset/8c8836e7dcdad528300a92be06f0f66e

[8] Huang, L. H.-C. (2022). *Applying social network theory and analysis* [Online resource]. ResearchGate. https: //doi.org/10.13140/RG.2.2.10425.36960

[9] Xinhua Net. (2025, November 18). *Investigation on AI-modified content: Primary students obsessed with "Foreign Shan Hai Jing"*. https: //www.news.cn/politics/20251118/14d36972c3ae405b89f981cd8eaff638/c.html

[10] Yuan, Y., & Fang, Z. Q. (2025, August 6). *Enhancing adolescent digital literacy requires collaborative efforts*. People's Daily.