“Gaming the System” on Xiaohongshu: Chinese University Graduates’ Responses to AIGC Detection Policies

Lixuan Wang, University of Macau (first author)

In Son Zeng, University of Macau (first author)

Ze Hong, University of Macau*(corresponding author)

Abstract

As Chinese universities increasingly use AI Generated Content (AIGC) detection scores to regulate graduation theses, students must navigate unclear policies, unreliable detection tools, and high-stakes degree requirements. This study examines how graduating students respond to such detection regimes through a mixed-methods analysis of Xiaohongshu, a major Chinese social media platform. We analyzed 191 posts and 6,025 comments related to AIGC detection and conducted semi-structured interviews with 18 recent graduates. Combining platform analytics, VADER-inspired sentiment profiling, LLM-assisted annotation, human validation, and interview-based thematic analysis, we show that AIGC detection is not merely a technical integrity procedure but a platform-mediated governance ecology. Students share score-reduction strategies, compare detectors, seek emotional reassurance, and sometimes enter gray-market rewriting services. Detection thresholds intended to control AI use may therefore stimulate more strategic paraphrasing, platform comparison, and detector-oriented writing. The findings suggest that universities should move beyond single-score thresholds toward transparent, educative, and appealable AI-use policies.

Keywords

AIGC detection, Academic integrity, Generative AI, Social media, Social computing, Higher education

1 Introduction

The rapid diffusion of generative AI (GenAI) in higher education has transformed academic writing into a site of institutional governance involving integrity, authorship, assessment, and responsible technology use [27, 30]. Across multiple national contexts, students now routinely use GenAI for brainstorming, language polishing, literature summarization, and even full-text drafting [4, 15]. A growing body of research has begun to address the academic integrity concerns in the AI age [6, 11]. Unlike translation software or ghostwriting services, GenAI occupies a gray area: AI-assisted work does not map neatly onto conventional definitions of plagiarism, yet it cannot be treated as unproblematic in terms of authorship and academic integrity [9, 10]. Universities, publishers, and policymakers have therefore faced a major governance challenge in defining acceptable use [12, 22, 30]. Strict prohibition may conflict with technological development and risk deepening educational inequity for students who rely on these tools to bridge language gaps, disabilities, or other access barriers [22, 23].

Student perceptions of GenAI in academic writing are neither uniformly accepting nor uniformly resistant. Students generally recognize the value of AIGC technologies for personalized support, writing assistance, and idea generation [18, 42], while also expressing concerns about accuracy, over-reliance, and ethical ambiguity [3, 10, 19]. Even among students who report gains in efficiency and creativity, anxiety remains over the reliability of domain-specific information and whether GenAI meaningfully improves academic performance [15]. This study therefore focuses not on the general merits and risks of AIGC, but on how these debates are experienced in students’ perceptions and behaviors. Prior research suggests that students’ uncertainty about acceptable use stems less from an intent to deceive than from institutional ambiguity [23]. Many students do not regard AI use in academic activities as an extension of conventional academic dishonesty, but rather as a distinct category of academic conduct/misconduct; acceptance levels are considerably higher for grammar checking and content summarization than for full-text generation [29], a spectrum of use that threshold-based detection models are structurally ill-equipped to capture. These perceptual differences are also marked by significant heterogeneity across disciplines, university levels and contexts of use. The faculty-student gap in perceptions of GenAI use tends, for instance, to be wider in the natural sciences than in the social sciences and humanities [19, 49]. In high-stakes academic writing contexts such as thesis writing, this uncertainty surfaces as cognitive dissonance. Interview-based studies conducted

in both China and the United States find that postgraduate students simultaneously acknowledge the practical value of GenAI and express uncertainty about where the boundary between legitimate use and plagiarism lies, as well as confusion about whether to disclose GenAI use in their work [20, 38]. This dissonance reflects a structural misalignment between institutional responses and students' actual needs.

Worldwide, scholarly and institutional attention has in parallel shifted toward how such use should be defined, regulated, and pedagogically framed. In China, this global policy challenge is especially pronounced in the context of graduation theses, where thesis completion is directly tied to degree conferral and students' overriding concern centers on whether they can pass their thesis defense and graduate successfully. The primary governance mechanism adopted across Chinese universities has been to partner with third-party providers to detect and quantify the proportion of AI-generated content (AIGC) in graduation theses, using a percentage score to regulate AI use by both undergraduate and postgraduate students. For example, Fuzhou University introduced a trial AI content detection system for undergraduate thesis review, while Tianjin University of Science and Technology warned students when AI-generated content exceeded 40% [51]. More recent reporting indicates that some universities have set stricter AIGC percentage limits, such as 20% for humanities theses and 15% for science, engineering, and medical theses, while others have adopted 40% as an upper threshold [48]. There are also some institutions that have developed more comprehensive frameworks. For example, Fudan University established a "six prohibitions" framework stating the permissible boundaries of AI-assisted writing in undergraduate theses, while simultaneously opening at least 100 "AI+ courses" in the 2024–2025 academic year, aimed at preserving originality and critical thinking while deterring future misconduct. However, balancing such restrictions against the broader potential of AI to support and democratize access to complex academic tasks remains a genuine challenge, therefore no consensus approach has emerged yet. Students in mainland China also face a distinctive set of constraints largely absent in other national contexts, for the leading generative AI tools such as ChatGPT, Gemini, and Claude lie outside the locally accessible digital environment [50]. These conditions give the broader tensions described above a particularly distinctive character within the Chinese higher education context. In practice, most institutions have settled on a single AIGC percentage threshold as their operative standard.

How this threshold-based approach shapes student cognition and behavior is therefore a central concern. Students' willingness to use AIGC tools is substantially influenced by how institutions define the legitimacy of such use [42]. Trust operates as a critical mediator in GenAI adoption [43], trust in tool accuracy and trust in procedural fairness together determine whether student engagement remains responsible or becomes purely instrumental. Research further suggests that when students seek clear policy guidance, the motivation is more often a desire to act responsibly within a defined framework than to carve out permissive space [23]. Institutional intervention, however, does not always produce its intended effects. When universities respond to the complexities of AI use through compliance-based detection rather than ethical cultivation, the intellectual development that higher education is meant to foster may be quietly undermined [17, 33]. Drawing on a dual teacher-student perspective to assess undergraduate thesis quality in China, Yan et al. [49] offer a particularly instructive set of findings: both the degree of GenAI involvement and students' perceived competence gains were consistently associated with higher thesis quality evaluations, while explicit institutional policies, though broadly beneficial, proved of limited effect. Path analysis further revealed that teachers tend to explain quality improvements through a positive coupling of "norms and participation," whereas students are more likely to translate "ability gains" into perceived quality. The uneven gap between faculty perceptions of students' AI skills and the actual deployment of those skills introduces additional bias into faculty judgments about student knowledge and the authenticity of assessed work [1]. This teacher-student divergence in AI-related cognition points to a relational gap that current AI governance practices in China have yet to adequately address.

The threshold-based detection arrangements described above are complicated in several respects. The first concerns the reliability of detection tools. Beyond widely documented accuracy limitations [45], leading AI detectors disproportionately flag non-native English writing as AI-generated [28], a bias with direct relevance to Chinese student populations. Threshold-based detection is also theoretically vulnerable to paraphrasing attacks, making reliable discrimination between human and AI text practically intractable [39]. When AIGC scores are nonetheless tied to degree eligibility, the problem extends beyond technical misclassification: students increasingly orient their work toward satisfying algorithmic criteria rather than addressing human evaluators who exercise disciplinary judgment, a governance dynamic consistent with Goodhart's Law, whereby a measure ceases to be a good measure once it becomes a target [41]. The second complication concerns the systemic behavioral responses this arrangement generates. When institutional thresholds are perceived as arbitrary or circumventable, and when students are dissatisfied with the teaching and learning environment, effort tends to be displaced from substantive learning toward compliance performance, such as contract cheating [7, 26]. In the AIGC context, this displacement is visible in insecure platform checks, paraphrasing strategies, and tensions generated by the opacity of existing

detection frameworks [39, 52]. Yet existing research has largely treated AI use and detection policy as separate objects of inquiry, and how AIGC detection thresholds generate a platform-mediated behavioral ecology of help-seeking, peer troubleshooting, emotional support, and grey-market service exchange remains underexamined.

This study therefore examines how graduating students in China respond to institutional AIGC detection policies through public discourse on Xiaohongshu, a major Chinese social media platform. Three research questions guide the inquiry:

RQ1: How do graduating students in China perceive and make sense of the AIGC detection thresholds set by their universities?

RQ2: What behavioral responses and coping practices do students develop on Xiaohongshu when navigating AIGC detection requirements?

RQ3: What does students' online discourse reveal about the gap between current detection-based governance and students' actual academic experiences?

The remainder of this paper is structured as follows: Section 2 details the data collection and analytical methodology. Section 3 presents the findings from both the computational analysis and the interviews. Section 4 discusses the theoretical and practical implications of these findings, and section 5 concludes with a summary of key contributions.

2 Methodology

This study employs a mixed-methods approach to investigate Chinese university graduates' responses when facing AIGC detection requirements for their graduation theses. We selected the Chinese social media platform Xiaohongshu (RedNote) as our primary research site. Evolving from an e-commerce and lifestyle app into a dynamic space for peer-generated content, Xiaohongshu now boasts over 339 million monthly active users, with 43% aged 18-24 and 36% aged 25-34 [37], often referred to as "Chinese Instagram". Xiaohongshu was selected because this age profile closely aligns with our target demographic of graduating students. Additionally, its users demonstrate high familiarity with AIGC tools [32], and it is increasingly recognized by researchers as a valuable site for studying students due to the active sharing of academic information and daily life experiences among young people [31, 46]. To analyze these digital traces, we conducted a textual analysis of publicly available posts and comments. In accordance with standard internet research ethics [2, 25], informed consent was not required for public data. However, to ensure user privacy and minimize interference, we strictly anonymized all personally identifiable information and institutional affiliations, exclusively analyzed historical data posted at least one year prior [35].

To complement the observational social media data and achieve methodological triangulation [24], we subsequently recruited 18 university students (2024 and 2025 cohorts) through Xiaohongshu for semi-structured interviews. These interviews were designed to deeply explore the internal logic driving their attitude formation, strategy selection, and emotional experiences regarding the AIGC detection policy. Finally, to ensure an accurate interpretation of both the interviews and online texts, the research team relied on their natural familiarity with the platform's discourse as active Chinese users.

2.1 Social Media Data

2.1.1 Data Collection. Data were gathered using Xiaohongshu's search functionality with relevant Chinese-language terms such as "毕业论文" (graduation thesis), "AIGC率" (AIGC rate), and "AIGC检测" (AIGC detection). These keywords were selected based on preliminary exploratory searches indicating their prevalence in discussions regarding the topic. For each included post, we collected basic attributes when applicable (e.g., title, body text, hashtags, publication timestamp, and IP location), multimedia links, and interaction metrics (e.g., counts of likes, shares, saves, and comments). Regarding comments, first-level comments were first collected, and then all subsequent replies were iteratively gathered as second-level comments. For both levels of comments, we extracted comprehensive metadata, including the commenter's user ID, publication timestamps, IP locations, and any embedded image links. We also captured comment-level engagement metrics, specifically the number of likes for all comments and the reply counts for first-level comments. The resulting dataset comprises all posts and their associated comments published under these keywords prior to January 5, 2026. The final analytical corpus consists of 191 posts and 6,025 comments, including 2,840 first-level comments and 3,185 second-level replies.

2.1.2 Data Analysis. In the exploratory data analysis part, we analyzed the hierarchical dataset by combining platform analytics, geographical, and temporal analysis. To systematically extract users' opinions, thoughts, and impressions regarding various topics, subjects, and services on Xiaohongshu platform, we employed a hybrid sentiment analysis approach [40, 44]. We first utilized a custom adaptation of VADER (Valence Aware Dictionary and sEntiment Reasoner) for sentiment profiling [21]. Given that social media discourse is replete with emotion-rich

and attitude-heavy expressions [40], this domain-specific Chinese VADER analysis first provided crucial heuristic inspiration in baseline affective landscape and slang usage, serving as a vital foundation that informed and refined our subsequent LLM coding framework.

First, we developed this customized VADER-inspired lexicon and rule-based sentiment analysis script tailored to Chinese social media text and the specific context of AIGC detection. We constructed different dictionaries assigning valence scores to Chinese slang and expressions. For example, the positive lexicon included terms like "过了" (I passed, +1.8) and "降到5%" (AIGC rate dropped to 5%, +2.3), while the negative lexicon captured academic distress with terms like "崩溃" (I collapsed, -2.6), "救命" (help me, -2.4), and "延毕" (delay graduation, -2.7). The algorithm also accounted for intensifiers (e.g., "太"(much) multiplying the score by 1.35) and punctuation emphasis (e.g., exclamation mark multiplying the score by 0.18). A compound sentiment score was calculated for each text using the following normalization formula:

$$compound = \frac{raw}{\sqrt{raw^2+15}}$$

In this equation, *raw* represents the sum of the valence scores of all matched sentiment words after applying modifier rules. The denominator normalizes this raw sum to a continuous bounded scale approaching -1 (most negative) to +1 (most positive), with 15 serving as a standard smoothing constant to approximate the maximum expected raw score [21]. Based on this continuous score, texts with compound scores ≥ 0.35 were classified as positive, ≤ -0.35 as negative, and those in between as neutral. Furthermore, a weakly supervised "ad-prediction" feature was integrated to identify and filter out soft advertisements and underground market promotions on the Xiaohongshu platform using a sigmoid-based confidence threshold (≥ 0.60).

Building upon the initial insights from our Chinese VADER profiling, we employed the DeepSeek-v4-flash model to assist with the automated coding process. This model was specifically chosen for its outstanding proficiency in Chinese natural language understanding and complex reasoning, offering capabilities comparable to leading closed-source models at a highly efficient computational cost [14]. To ensure the rigor of the automated coding process, we designed a hybrid prompting strategy. Structurally, we adopted the CO-STAR framework (Context, Objective, Style, Tone, Audience, Response) to define the study's background, objectives, broader context, and output constraints. Cognitively, we integrated Few-Shot Prompting with a structured reasoning pathway [8]. Specifically, building upon the principles of Chain-of-Thought reasoning [47], we explicitly embedded logical execution sequences and tie-breaking rules into the prompt. This guided the LLM to evaluate the text step-by-step according to strict predefined criteria before assigning a final code. Through this systematic prompting strategy, we operationalized a four-dimensional coding scheme tailored to the platform's unique discourse (the complete prompts are provided in the Supplemental Information):

- **Methods_and_Strategies**: Identifies AIGC detection methods and AI-reduction strategies shared by users. We defined six categories: Manual_Modification, Translation_Method, AI_Prompting, Tool_Recommendation, Purchased_Service_Experience, and None.
- **Supply_and_Demand**: Captures underground market dynamics related to AI-reduction services. Three categories were defined: Supply_Offering (users offering paid/free services, including soft advertisements), Demand_Seeking (users requesting help, tools, or resources) and None.
- **Attitudes**: Measures users' macro-level critiques targeting different stakeholders in the AIGC detection ecosystem. Four categories were defined: Critique_Platforms (targeting detection platforms), Critique_Policies (targeting university policies), Critique_Tech (targeting AI detection technology itself), and None.
- **Emotion**: Assesses users' psychological states during the thesis AIGC detection process. Five categories were defined: Anxiety_Panic, Frustration_Exhaustion, Confusion, Relief_Joy, and None.

To ensure reliability, we first conducted a trial run and the Inter-Rater Reliability (IRR) test. A random sample of 300 comments was first independently coded by two trained human coders to establish a "ground truth" through consensus on discrepancies, and DeepSeek-v4-flash then processed the same sample. We calculated Fleiss's Kappa coefficients to evaluate the agreement between the LLM coder and human coders. Results indicated that DeepSeek-v4-flash achieved satisfactory reliability across all four dimensions (Overall Fleiss's Kappa > 0.60) (see Supplemental Information). Consequently, the model was deployed to code the full comments dataset.

2.2 Interview Data

2.2.1 Participants. Participants were recruited through purposive sampling on Xiaohongshu communities. We selected 18 respondents who met both of the two inclusion criteria: (1) graduated from a Chinese higher education institution in 2024 or 2025; (2) required to submit a graduation thesis as core assessment material and experienced

or encountered school requirements for AIGC detection. Among them, 13 participants were undergraduate graduates and 5 were postgraduate graduates.

We conducted semi-structured interviews lasting 30-45 minutes per person, covering the following topics: (1) students' daily AIGC usage, including commonly used LLM models, frequency, and specific applications in their graduation theses; (2) institutional policies regarding AIGC use in graduation theses and their implementation, as well as students' attitudes toward these regulations; (3) students' actual experiences with and trust in AIGC detection technologies and platforms; and (4) the specific methods students adopted to reduce AIGC scores, along with their emotional responses to this process.

Questions were open-ended to allow for emergent themes. All 18 interviews were audio-recorded with consent, transcribed using Lark.ai and Tencent Meeting followed by manual verification, and then translated into English for further coding and analysis. To maintain confidentiality, all interview data will be securely stored on a local password-protected hard drive of the authors for a period of five years. Participants received a small incentive of 15-25 RMB for their time. Participant demographics are detailed in Table 1, and complete interview protocols are available in Supplemental Information.

Participant	**Age**	**Gender**	**Major**	**Graduation Year**
U1	22	Female	E-commerce	2025
U2	22	Female	Auditing	2025
U3	21	Female	Economics	2025
U4	22	Male	Computer Science	2025
U5	22	Female	Economic Management	2025
U6	23	Male	New Energy	2024
U7	22	Female	English	2025
U8	22	Female	Management of Cultural Industry	2025
U9	22	Female	Mechanical design	2025
U10	22	Male	Physics	2025
U11	23	Female	Russian	2025
U12	21	Female	Finance	2025
U13	21	Female	Social Science	2025
P1	25	Female	History	2025
P2	29	Female	Art Design	2024
P3	26	Female	Management Science and Engineering	2025
P4	26	Female	Law	2025
P5	27	Female	Sociology	2025

Table 1. Demographic Information about Participants

2.2.2 Interview Analysis. We analyzed the interview data using a hybrid approach that integrates qualitative thematic analysis with computational network modeling. First, researchers conducted close readings of the transcripts to identify recurring concepts. We then developed a structured codebook covering key dimensions, including institutional policies, AIGC usage scenarios, risk perceptions, and emotional impacts. To process the text systematically, we applied a Python script for preliminary text chunking and keyword-based code assignment. Researchers subsequently reviewed and manually adjusted all automated codes to ensure contextual accuracy. This human-in-the-loop process maintained qualitative rigor while improving analytical efficiency.

Following the coding phase, we employed computational network analysis to map the structural relationships among the identified themes. Using the NetworkX library, we constructed a theme co-occurrence matrix to calculate how frequently different codes appeared together within individual participant narratives. This matrix allowed us to generate network graphs and hierarchical tree structures. These computational models revealed macro-level linkages between themes, such as the underlying connections between institutional uncertainty and student anxiety. In the results section, we triangulate these network patterns with direct interview quotes to provide a detailed and empirically supported account of the students' experiences.

3 Results and Findings

The final corpus contains 6,216 Xiaohongshu texts organized into three conversational layers: 191 original posts, 2,840 first-level comments, and 3,185 second-level comments. This hierarchy is analytically important because the three layers serve different communicative functions. Original posts usually frame AIGC detection as an urgent graduation-thesis problem; first-level comments transform these posts into requests, advice, emotional responses, or service offers; second-level comments extend the discussion into micro-level troubleshooting and interaction. Therefore, the dataset should be interpreted as a platform-mediated interaction rather than as a flat collection of independent opinions. To strengthen the credibility of the computational annotation, the study used human interpretation as the reliability anchor. The LLM-generated labels were therefore used for large-scale pattern discovery, while interpretation remained grounded in a human-validated coding framework. This combination is important because the texts are short, emotionally charged, and often ambiguous.

3.1 Social media discussion dispersion: geography, timing, engagement and lag

The geographic visualizations show that AIGC detection discussion is China-centered but not locally confined. Original posts are the least geographically complete and should be interpreted cautiously since only 50 of 191 posts contain known IP locations. Among known post locations, Figure 1 shows that Jiangsu appears most frequently, followed by Guangdong and Beijing. In contrast, comment-level geography is much more complete: 2,764 of 2,840 first-level comments and 3,086 of 3,185 second-level comments contain known locations. Both comment layers are led by Guangdong, with 385 first-level comments and 383 second-level comments. The high normalized entropy values for first-level and second-level comment locations (.829 and .832) indicate that the discussion is not restricted to one university, city, or province. Instead, AIGC detection anxiety circulates across a distributed student population, with stronger visibility in digitally active and higher-education-dense regions such as Guangdong, Jiangsu, Hubei, Beijing, and Sichuan.

The post map shows where original public problem narratives are initiated, while the first-level and following-comment maps show where advice seeking, emotional response, and troubleshooting are amplified. The denser distribution in the comment maps suggests that once a post becomes visible, it attracts participation from a broader audience than the original poster group. Overseas locations appear only marginally, so the core interaction remains domestic and China-centered, with overseas engagement functioning as an extension rather than the main discussion base.

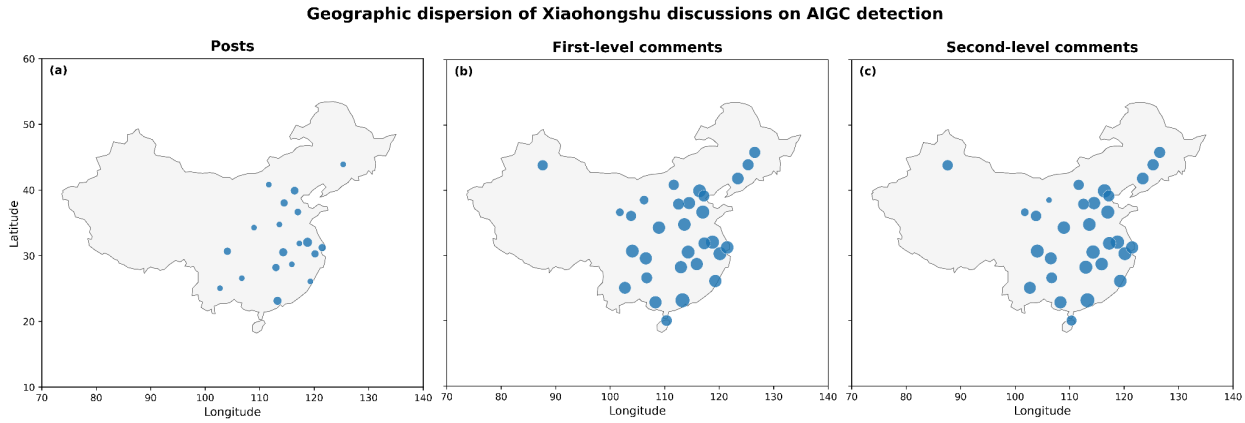

Figure 1. China map of post and comment locations

The time-analysis figures (Figure 2, Figure 3) show that AIGC detection discourse is deadline-sensitive. Posts range from April 12, 2024 to January 5, 2026; first-level comments range from April 15, 2024 to January 5, 2026; and second-level comments range from April 18, 2024 to January 5, 2026. Activity peaks around April-May and December-January, corresponding plausibly to thesis revision, checking, submission, and final review procedures. Posts peak in December 2025 (46 posts), April 2025 (39), and May 2025 (24). First-level comments peak in December 2025 (715), May 2025 (605), and January 2026 (452), while second-level comments peak in May 2025 (744), December 2025 (721), and January 2026 (416). The comment layers therefore often amplify or prolong the post layer, demonstrating that Xiaohongshu threads become temporary consultation spaces when students face urgent detection deadlines.

Lag analysis further shows how Xiaohongshu threads function as short-cycle consultation spaces. The median lag is four days for first-level comments and seven days for second-level comments. About 60.4% of non-negative first-level comments and 50.5% of non-negative second-level comments occur within seven days of the original post. The platform thus operates as a short-cycle peer-support infrastructure: posts remain most useful during the first week, but highly searchable advice can be reactivated when later students encounter the same institutional threshold or detector problem.

The engagement structure is highly unequal. The median post receives only three comments, nine likes, two shares, and five saves, yet rank-size patterns show that a small number of posts receive disproportionate attention. The posts that become visible reference points tend to be practical rather than theoretical: they share tool comparisons, successful score-reduction experiences, platform warnings, or step-by-step procedures. First-level comments sometimes become secondary anchors of the conversation. In 35 posts, at least one first-level comment reached 50% of the original post likes, showing that peer testimony can become as influential as the initial post. Second-level comments rarely exceed post-level visibility, but they are valuable for narrow troubleshooting.

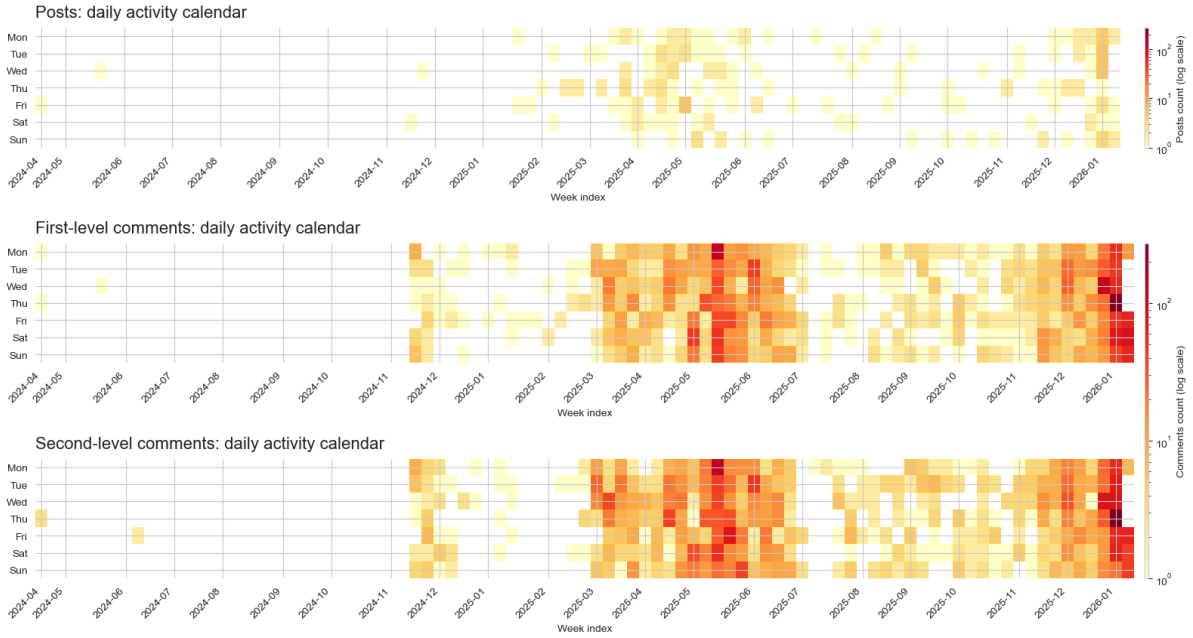

Figure 2. Daily activity calendar for posts, first-level comments, and second-level replies. The heatmap supports the interpretation that AIGC detection conversations intensify in deadline-sensitive periods.

The word-cloud visualizations explain why certain posts and comments become useful. Across posts and comments, the dominant terms include “AI,” “AIGC,” “检测” (detection), “查重” (similarity/plagiarism check), “论文” (thesis/paper), “结果” (result), “知网” (CNKI), “PaperPass,” “降重” (score reduction), “避雷” (avoid/blacklist), and “查出来” (detected by the system). This vocabulary is operational and comparative rather than philosophical. Users are not mainly debating whether AI should exist in education; they are asking which detector is closest to the university system, how to reduce scores, whether a result will pass, and which tools or services should be avoided. The visual evidence therefore supports the claim that AIGC detection pushes students toward detector-oriented compliance and platform comparison.

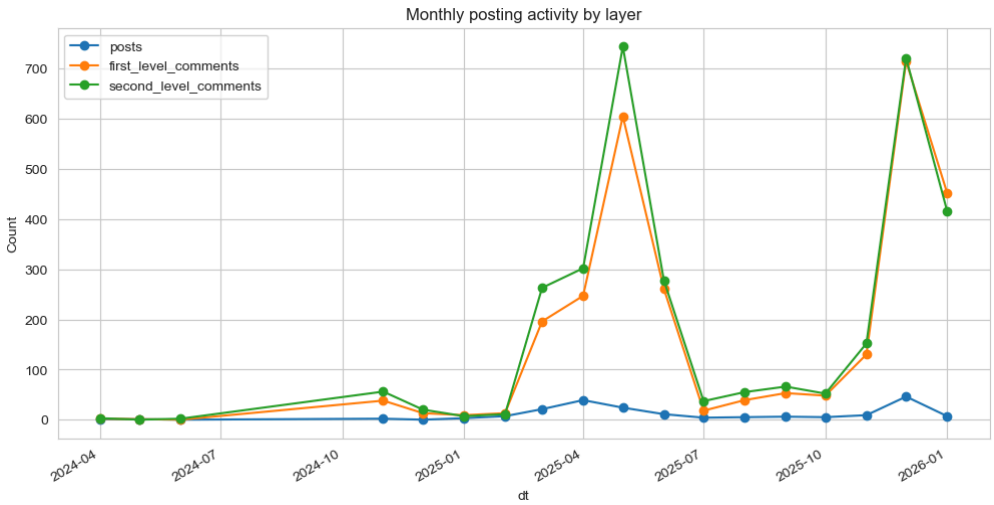

Figure 3. Monthly posting activity by layer. The peaks in comments and replies demonstrate that response activity can exceed post creation during thesis-checking periods.

3.2 Lexical affect and ad-confidence distributions

Figure 4 reports lexical sentiment compound scores and advertisement-confidence scores across three conversational layers. The lexical compound plots show that most Xiaohongshu texts cluster around zero, indicating that a large share of posts and comments are informational, procedural, or too brief to carry strong lexical sentiment. This is particularly visible in first-level and second-level comments, where short requests, platform names, score reports, and troubleshooting exchanges often appear as lexically neutral. By contrast, original posts show wider affective dispersion, with more visible positive and negative tails. This suggests that posts are more likely to frame the issue through success stories, warnings, frustration, urgency, or platform criticism.

The ad-confidence distributions show that most comments receive low advertisement-confidence scores, indicating that general discussion dominates the corpus. However, all three layers contain right-tail cases with higher ad-confidence, with the post layer being the most dispersed. This suggests that promotional or service-oriented content is not the dominant form of discussion, but they are visibly embedded within the wider help-seeking ecology. These baseline distributions justify the need to combine lexical profiling with human-validated annotation, because slang, emoji-rich promotional texts may not be fully interpretable through sentiment scores alone.

3.3 Strategy, market exchange, emotion and critique

The annotation results show that students respond to AIGC detection through shifting strategies rather than a single coping pattern. In original posts, Manual_Modification is the most frequently recommended strategy, appearing 51 times and accounting for 77.27% of strategy-coded posts. Successful posters often present manual rewriting as the safest and most legitimate way to reduce AIGC scores. This strategy is frequently associated with Relief_Joy because posts often describe an already achieved outcome: the author lowered a score, passed a check, or found a workable procedure. However, when commenters attempt to reproduce manual modification, the emotional meaning of the strategy changes. Manual rewriting becomes associated with repeated checking, inconsistent reports, and deadline pressure. What appears in posts as an effective route to success becomes, in the comment layers, a source of exhaustion. This emotional inversion helps explain the strategic shift from manual modification to automation in deeper replies. For instance, AI_Prompting rises from 4.55% in posts (N = 3) to 26.92% in second-level comments (N = 84). This poses one of the most important findings: policies intended to control AI-generated writing can paradoxically stimulate more AI-mediated rewriting, prompting, paraphrasing, and detector-optimization behavior. Our interviewees narrated this shift directly. One interviewee (U7) described feeding flagged passages back into the same AI tool: “Once the system marked something red, I’d send that section back to ChatGPT and ask it to rephrase it.” U11 also made the same decision after exhausting manual alternatives: “In the end the goal was just to graduate. If it could actually bring the score down, using AI to do it was just fine.” What begins as an effort to authenticate the text ends, at the limit of exhaustion, as a decision to outsource the problem back to the machine.

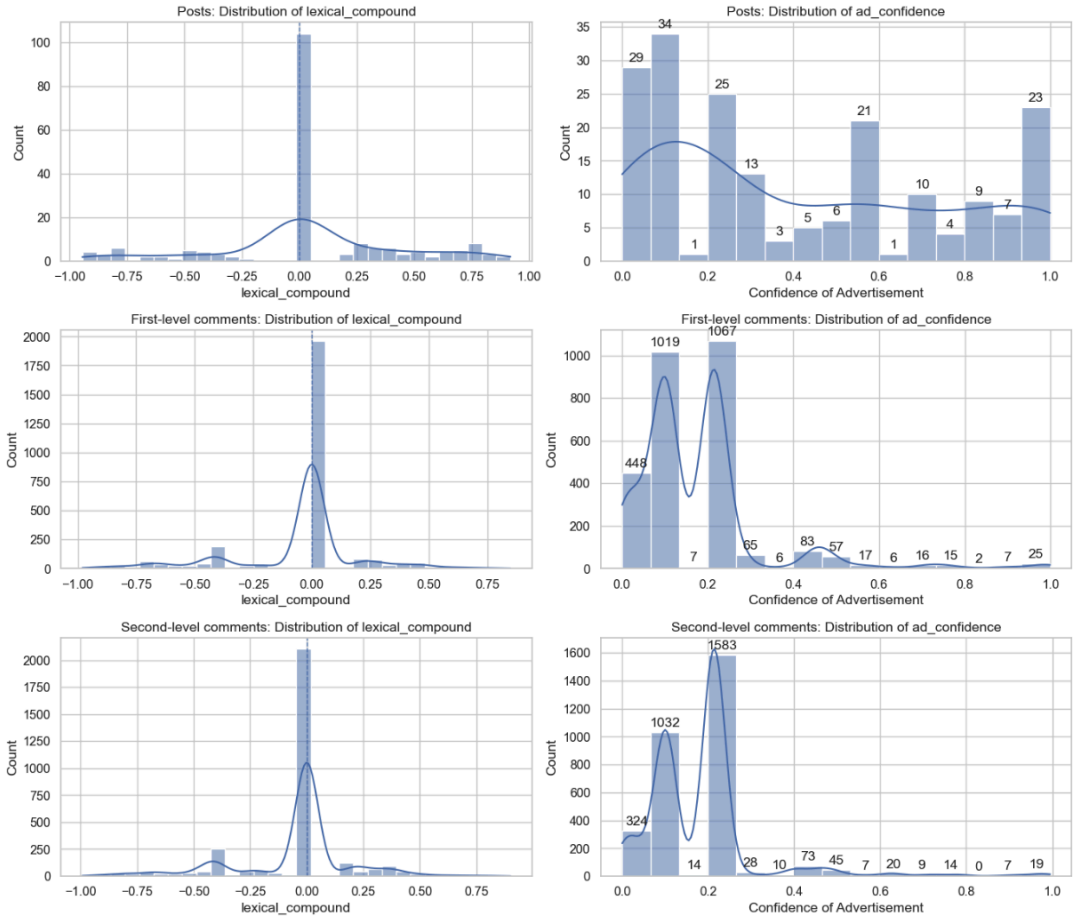

Figure 4. Distribution of lexical sentiment compound scores and advertisement predictions for posts, first-level comments and second-level comments.

The supply-and-demand annotations reveal a sharp imbalance between users seeking resources and users offering services. Demand-seeking comments greatly outnumber supply offerings, creating a baseline condition of perceived resource scarcity. Students ask for detector names, score-reduction methods, fees, service contacts, and evidence that a method has worked. Supply offerings appear in the same peer-support space, often using informal peer-like language rather than conventional advertising. Among first-level comments, 71 were flagged as advertisements, while 2,769 were classified as general discussion. High-confidence advertisements commonly promised "AI-rate reduction," "manual service," "reasonable price," "tested effective," or "after-sales support." This supports the paper's grey-market argument: the market is not external to the discourse, but embedded within the same platform affordances that help anxious students seek advice.

Interviews show that students often enter these markets after repeated failed attempts at legitimate self-editing. U9 explained, "The reason I stopped doing it myself was that I was already going crazy. I'd rather pay and just hand over the responsibility." When asked about the effectiveness of such services, she said: "...the results were okay, though one failed attempt had burned a paid detection check in the process. I let it go because the provider agreed to split that extra cost with me." P4 described paying 35 RMB for a Xiaohongshu service provider, only to receive an AIGC detection score that barely changed: "I messaged him afterwards but he just stopped responding. That's when I realised it was a scam." Both cases suggest that grey-market participation is driven less by calculated misconduct than by emotional depletion. Without contracts, verification mechanisms, or recourse, students are not buying a reliable solution; they are temporarily transferring an unbearable burden to someone else, with no assurance that the solution will work.

The emotions these individual accounts show are not isolated; they recur at scale across the annotated dataset. The VADER-informed and LLM-assisted annotations all show a strong affective transition across discourse layers. Original posts often narrate successful outcomes and therefore contain a comparatively high share of

Relief_Joy. In the post layer, Relief_Joy appears 41 times, accounting for 41.41% of emotionally annotated posts. This does not mean the discourse is positive overall. Rather, successful posts become visible because they provide hope and procedures for others. Once users enter first-level comments, the emotional center shifts sharply. First-level comments operate as a sink for collective burnout: Frustration_Exhaustion becomes the dominant emotion (N = 584, 42.88%), followed by Anxiety_Panic (N = 348, 25.55%). As one interviewee (P1) said: "I felt like I was going crazy…those months writing the thesis, my head was about to explode. I only wanted to pass the thesis defense and that was it." What the computational data registers as Frustration_Exhaustion is, in these accounts, a process of progressive detachment from the thesis as an object of intellectual investment.

Attitudinal coding shows this exhaustion. Across all three layers, Critique_Platforms dominates the attitudinal landscape: 61.11% of attitudinally annotated posts (N = 33), 59.25% of first-level comments (N = 477), and 57.96% of second-level comments (N = 455) criticize commercial detection platforms. The fused attitude-emotion distribution shows that the densest cell is the intersection of Frustration_Exhaustion and Critique_Platforms. Students repeatedly frame detection platforms as inaccurate, opaque, expensive, and "money-grabbing." This pattern indicates that burnout is not interpreted merely as ordinary academic stress. Students attribute exhaustion to a commercial detection ecosystem that they believe profits from uncertainty. Interviewees identified specific dimensions of this critique that attitudinal coding alone cannot capture. P3 described querying the same document across multiple detection platforms and receiving results ranging from 20% to 90%: "I genuinely couldn't tell which one to believe." U11 identified what she saw as a core structural contradiction: "The system uses AI to judge whether something was written by AI, and then the paid service that's supposed to lower your AIGC rate also uses AI to rewrite it. Isn't that completely self-contradictory?" The critique of platforms is therefore not simply a complaint about price; it is a challenge to the epistemic authority claimed by systems that students experience as opaque, inconsistent, and internally contradictory.

3.4 Interview triangulation

The interview visualizations triangulate these computational findings. The most frequent interview codes are AIGC polishing (23), policy ambiguity (22), policy threshold governance (17), informal detector market (16), AIGC literature/data use (16), hallucination/accuracy risk (15), ethics guidance (15), and AIGC ideation/structure (15). These codes show that students experience detection not simply as a technical checking step but as a governance environment. Students often use AI for language polishing, structure ideation, and research assistance, yet they are unsure where legitimate assistance ends and misconduct begins. Numeric thresholds transform writing into score management, and uncertain official channels increase dependence on informal detectors and peer-shared advice.

The interview co-occurrence matrix (Figure 5) further links policy ambiguity, threshold governance, official and informal detectors, psychological pressure, and adversarial gaming. A tree-based analysis similarly groups these themes under broader domains of institutional context, writing-use scenarios, risk and trust judgments, student coping mechanisms, and learning or emotional consequences. Together, these analyses show that students' experiences of AIGC detection are distributed across institutional rules, tool reliability, writing practices, informal markets, and emotional pressure, rather than being confined to a single issue of AI use or academic misconduct.

4 Discussion

4.1 AIGC detection as a social computing issue

The results reposition AIGC detection from an academic integrity procedure to a social computing problem. Detection does not operate only at the moment when a thesis is checked; it becomes socially meaningful through platform interaction, peer advice, commercial service signals, and emotionally charged troubleshooting. Its consequences are distributed across multiple actors and infrastructures: universities define thresholds, detection platforms produce uncertain scores, social media users circulate coping strategies, and students interpret these signals under graduation pressure. Together, these interactions create a self-reinforcing cycle. Institutional requirements and ambiguous detection standards prompt students to seek information and reassurance online. Exposure to detector scores, success stories, and cautionary accounts heightens concern about potential risks, leading students to engage in repeated self-checking, rewriting, and comparison across platforms. As more students share their experiences and strategies, the volume of platform content grows, further shaping collective perceptions of detector reliability and risk. In this way, individual responses to uncertainty become public resources that reproduce and amplify the very concerns that motivated them. Within this ecology, institutional uncertainty fuels detector anxiety; detector anxiety encourages self-checking, rewriting, platform comparison, and workaround strategies; and these strategies generate perceived efficiency while also increasing dependence, market exposure,

and psychological pressure. The social computing contribution of this study is therefore to show how an institutional technology becomes publicly negotiated, emotionally amplified, strategically adapted, and commercially mediated once it enters everyday platform discourse.

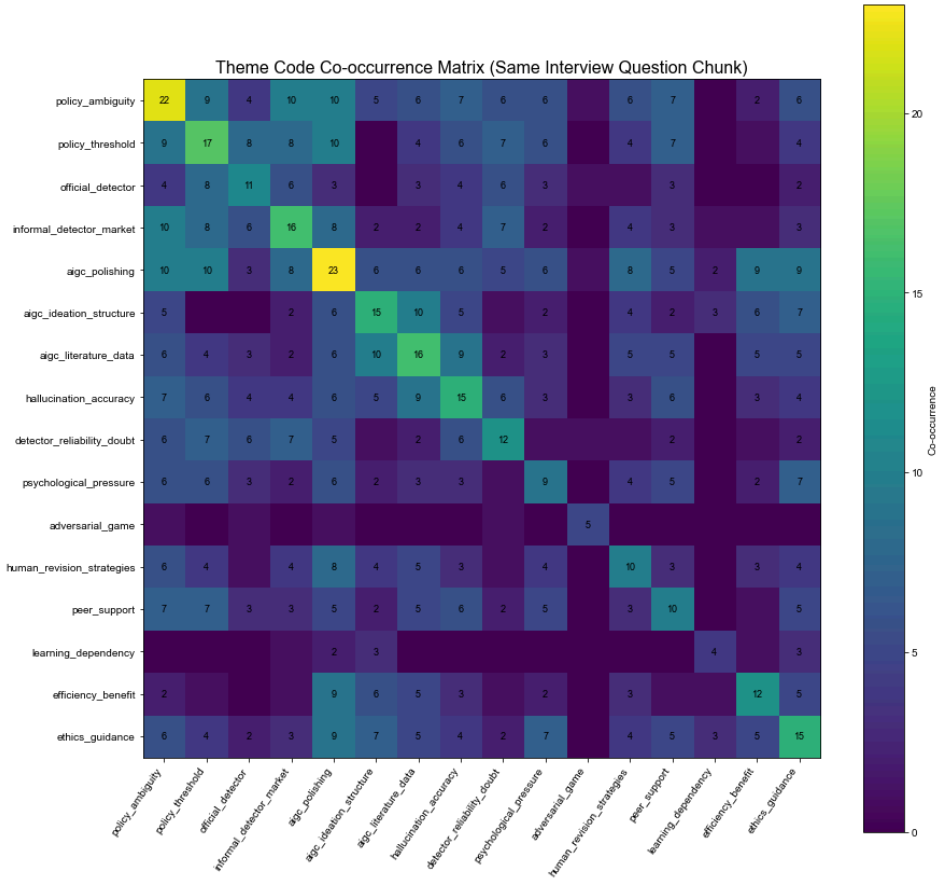

Figure 5. Interview theme co-occurrence matrix. Co-occurrences help locate the structural relationship between policy ambiguity, threshold governance, market dependence, and emotional pressure.

4.2 Algorithmic governance and tactical adaptation

The detection regime's authority does not derive from technical reliability, it derives from institutional embeddedness. In audit systems, normative force is gained through consequential use rather than accuracy, and in doing so the measure is transformed into the object of management itself [36, 41]. This dynamic is intensified in algorithmic systems, whose social power accrues precisely from their inaccessibility to those subject to them [5]. In this study, students cannot contest the logic of detection. They can only manage its outputs. This asymmetry instantiates that the detection regime does not merely identify pre-existing misconduct but generates the behaviors it was designed to regulate, a dynamic extensively documented in audit culture scholarship and the critical sociology of technology [5, 36, 41]. The vocabulary that saturates the dataset, terms like "AIGC rate," "CNKI threshold," "score reduction," and "pass," reflects a writing process reorganized around machine classification rather than scholarly communication. Weber-Wulff et al. and Liang et al. have documented the technical unreliability underlying this regime [28, 45], including systematic bias against non-native writers. Yet institutional authority remains intact. Students have no formal recourse when a score is ambiguous or a submission fails without explanation. They quickly adapt instead.

De Certeau's distinction between strategy and tactics offers a productive lens for understanding this adaptation [13]. Strategy belongs to those who hold a fixed institutional position from which they can define and enforce order. Tactics belong to those who lack that position and must maneuver within spaces they cannot redesign. Universities and detection platforms occupy the strategic position in this ecology. They set thresholds, define violations, and produce scores that carry academic consequences. Students act tactically within this arrangement. Manual modification, translation loops, cross-platform comparison, and AI-assisted paraphrasing are the

characteristic moves of agents who cannot change the rules but must navigate them, a form of improvised rationality in high-stakes institutional contexts [13]. This is what De Certeau calls the "weapon of the weak," and it fits the present situation closely. For most students in this study, the primary goal is not to harm academic integrity but to graduate. Thesis writing has become less a scholarly endeavor than a continuous and quiet navigation of a system controlled by others. What makes this tactical behavior significant in social computing terms is its collective character. Xiaohongshu functions as a shared tactical repertoire, where students document outcomes, circulate procedures, and refine strategies through peer interaction. Procedural sharing dominates the discourse because students recognize their shared structural position and its limits. The attitudes embedded in these practices are not straightforwardly oppositional. Students consistently frame their strategies as rational responses to an unreliable system rather than as violations, a reframing that Chan identifies in AI-related academic integrity discourse more broadly [9].

4.3 Affective pathways into the paid services

The pathway into grey-market services is better understood as affective than calculative. Contract cheating scholarship has largely framed outsourcing through rational cost-benefit models [7, 26], but what the data in this study describe is different. Grey-market participation appears consistently at the endpoint of exhausted legitimate strategies, after score-reduction attempts fail to stabilize, after translation loops degrade text quality, and after AI paraphrasing cycles back to flagged outputs. Newton argues that markets for outsourced academic work are sustained by structural conditions that make legitimate alternatives exhausting or unavailable [34]. The present findings add an affective dimension to this account. It is the emotional cost of having already tried, rather than a prior disposition toward misconduct, that makes students susceptible to commercial offers. What participants describe is burden transfer under depletion, not strategic investment. The emotional data across discourse layers show how this depletion is socially produced. Relief and joy dominate original post emotions at 41.41%, and these successful accounts raise implicit expectations for readers who have not yet found a solution. In the comment layer, the dominant emotions shift to frustration and exhaustion at 42.88% and anxiety and panic at 25.55%, together accounting for over two-thirds of first-level comment emotions. Ferrara and Yang demonstrate that emotional contagion on social media operates through exposure asymmetries of this kind [16]. Optimistic outcomes transmitted downward from successful posts elevate expectations, while negative affect circulates laterally through the comment stream, producing a net emotional ecology that maximizes vulnerability precisely when commercial offers appear in the platform flow. The stable presence of critique directed at detection platforms across all discourse layers, in original posts, first-level comments, and second-level comments alike, confirms that this alienation is structural rather than reactive. Students are not frustrated with individual outcomes. They are estranged from their own writing by a regime that has reduced scholarly work to a classification problem. A governance response that targets the grey-market transaction without addressing the upstream conditions of policy ambiguity, algorithmic opacity, and absent institutional support will, at best, displace this cycle. It will not interrupt it.

4.4 Implications, limitations, and future directions

Universities should respond to AIGC detection as an instructional design and governance problem, not only as a compliance problem. First, institutions should publish clear written rules distinguishing different forms of AI assistance, including brainstorming, grammar polishing, translation support, literature searching, data analysis assistance, and full-text generation. Second, they should avoid relying on a single numeric AIGC threshold without explaining false positives, uncertainty, appeal procedures, and the role of human review. Third, supervisors should discuss AI use early in the thesis process rather than introducing detection only at the final checking stage. Fourth, universities should provide official or institutionally vetted support for interpreting detection reports so that students do not have to rely on commercial public accounts, informal vendors, or peer rumors during high-pressure deadline periods. Finally, platforms should consider how service advertisements blend into peer-support threads and how vulnerable student demand can be exploited when detection anxiety is high. A constructive alternative is to move from a detection-only model toward a tiered and transparent AI-use policy. One possible "Wildcard" design would allow students to use AI for several designated assignments in a semester, even in courses where AI use is otherwise restricted, on the condition that they disclose the use, submit prompts and interaction methods, explain what they used AI for, identify what they revised or rejected, and reflect on how the process aided learning. This type of policy turns AI use into a transparent and accessible learning process. By requiring disclosure with presentation, a Wildcard design could weaken the incentive to purchase hidden rewriting services, and help students develop accountable AI literacy rather than merely learning how to evade detection.

Methodologically, the study demonstrates the value of combining human annotation, Fleiss's kappa validation, LLM-assisted coding, VADER-inspired affective profiling, platform analytics, and qualitative interviews

for studying fast-moving social media debates around AI-mediated educational technology. This mixed workflow operates at two levels. At the macro level, platform analytics, geographic patterns, temporal activity, and annotation distributions identify large-scale patterns in how AIGC detection is publicly discussed. At the interpretive level, human validation and interviews provide contextual grounding for understanding emotionally charged and slang-rich Xiaohongshu discourse, where anxiety, frustration, relief, commercial promotion, and peer support may appear in short or ambiguous texts. The methodological contribution is therefore not that any single technique fully captures student experience, but that triangulation can connect large-scale social media signals with human-centered interpretation.

This study has several limitations that point to future research directions. First, Xiaohongshu users are not representative of all Chinese university graduates. The platform may overrepresent students who are anxious, strategically active, commercially targeted, or willing to seek help publicly. Future research could conduct stratified surveys to test whether the observed patterns in Xiaohongshu generalize beyond visible social media users. Second, post-level IP-location data are incomplete, and even available IP locations should be interpreted as platform-provided indicators rather than precise geographic measures. Therefore, geographic analysis is descriptive in this study; it shows the diffusion and concentration of discussion rather than the true regional prevalence of AIGC detection anxiety and help seeking. Future research could combine social media data with institutional policy databases, regional university distributions, and official thesis-submission timelines to examine whether geographic peaks correspond to local policy differences, university density, or discipline-specific thesis requirements. Third, VADER-inspired lexicon compound scores and LLM-assisted annotations are interpretive aids rather than direct measurements of students' psychological states or intentions. Although human annotation and Fleiss's kappa validation strengthen the reliability of the coding process, computational labels still depend on lexicon design, prompt wording, category definitions, and validation sample size. Future research could expand the validation set, conduct error analysis across edge cases, and test whether categories such as Anxiety_Panic, Frustration_Exhaustion, Critique_Platforms, and Supply_Offering remain stable across time and platforms. Fourth, the present study focuses primarily on student discourse. Future research should include additional stakeholders, especially thesis committees, professors, detection-platform developers, and commercial service providers. With multi-stakeholder perspectives, the future work could clarify how institutional intentions, platform business models, instructor practices, and student interpretations become misaligned. It would also help evaluate whether alternatives such as tiered AI-use policies, transparent AI-use declarations, or "Wildcard" assignments can reduce grey-market dependence and adversarial detector gaming.

Taken together, these limitations clarify the scope of the present evidence. This study does not estimate the prevalence of AIGC-related misconduct among Chinese graduates; rather, it documents how AIGC detection is publicly negotiated within a visible social media ecology shaped by institutional ambiguity, commercial detection systems, peer support, emotional pressure, and strategic adaptation. Future research should extend this work through more representative sampling, strengthened annotation validation, policy-data linkage, and multi-stakeholder studies of alternatives to detection-centered governance.

5 Conclusion

This study examined how Chinese university graduates respond to institutional AIGC detection policies through Xiaohongshu discourse and interviews with 18 graduating students. Analysis of 6,216 posts and comments shows that policy ambiguity, detector unreliability, and graduation pressure push students toward score-management practices, including translation loops, repeated platform checking, AI-assisted paraphrasing, and grey-market services. In this sense, AIGC detection does not simply identify misconduct; it can also produce the behaviors it seeks to regulate by reorganizing thesis writing around algorithmic thresholds rather than scholarly communication. Students collectively develop workarounds on Xiaohongshu, while commercial actors enter the same peer-support spaces. These findings suggest that AIGC detection should be treated as a social computing and governance design problem, rather than a compliance tool. Universities should move beyond stricter detection toward clearer rules, human review, meaningful appeal mechanisms, and policies that frame constructive and transparent AI use as part of academic learning.

Acknowledgements

We sincerely thank all undergraduate and postgraduate participants who took part in the semi-structured interviews. Their reflections were invaluable for understanding how students perceive and respond to AIGC detection policies in graduation-thesis writing. We are grateful for their time, openness, and trust, and all participant identities have been anonymized to protect confidentiality.

References

[1] Alharbi, W.: Mind the gap, please!: Addressing the mismatch between teacher awareness and student AI adoption in higher education. International Journal of Computer-Assisted Language Learning and Teaching 14(1), 1–28 (2024). https://doi.org/10.4018/IJCALLT.351245

[2] Association of Internet Researchers: Internet Research: Ethical Guidelines 3.0. Association of Internet Researchers (2020). https://aoir.org/reports/ethics3.pdf. Last accessed 1 Jun 2026

[3] Baidoo-Anu, D., Asamoah, D., Amoako, I., et al.: Exploring student perspectives on generative artificial intelligence in higher education learning. Discover Education 3, 98 (2024). https://doi.org/10.1007/s44217-024-00173-z

[4] Barrot, J.S.: Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing 57, 100745 (2023). https://doi.org/10.1016/j.asw.2023.100745

[5] Beer, D.: The social power of algorithms. Information, Communication & Society 20(1), 1–13 (2017). https://doi.org/10.1080/1369118X.2016.1216147

[6] Bittle, K., El-Gayar, O.: Generative AI and academic integrity in higher education: A systematic review and research agenda. Information 16(4), 296 (2025). https://doi.org/10.3390/info16040296

[7] Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., van Haeringen, K., Rozenberg, P.: Contract cheating and assessment design: Exploring the relationship. Assessment & Evaluation in Higher Education 44(5), 676–691 (2019). https://doi.org/10.1080/02602938.2018.1527892

[8] Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al.: Language models are few-shot learners. In: Advances in Neural Information Processing Systems, vol. 33, pp. 1877–1901 (2020)

[9] Chan, C.K.Y.: Students' perceptions of "AI-giarism": Investigating changes in understandings of academic misconduct. Education and Information Technologies 30, 8087–8108 (2025). https://doi.org/10.1007/s10639-024-13151-7

[10] Chan, C.K.Y., Hu, W.: Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education 20, 43 (2023). https://doi.org/10.1186/s41239-023-00411-8

[11] Crompton, H., Burke, D.: Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education 20, 22 (2023). https://doi.org/10.1186/s41239-023-00392-8

[12] Dabis, A., Csáki, C.: AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative AI. Humanities and Social Sciences Communications 11, 1006 (2024). https://doi.org/10.1057/s41599-024-03526-z

[13] De Certeau, M.: The Practice of Everyday Life. University of California Press, Berkeley (1984)

[14] DeepSeek-AI: DeepSeek-V4-Flash. Hugging Face model card (2026). https://huggingface.co/deepseek-ai/DeepSeek-V4-Flash. Last accessed 1 Jun 2026

[15] Fan, L., Deng, K., Liu, F.: Educational impacts of generative artificial intelligence on learning and performance of engineering students in China. Scientific Reports 15, 26521 (2025). https://doi.org/10.1038/s41598-025-06930-w

[16] Ferrara, E., Yang, Z.: Measuring emotional contagion in social media. PLOS ONE 10(11), e0142390 (2015). https://doi.org/10.1371/journal.pone.0142390

[17] Gallent Torres, C., Zapata-González, A., Ortego-Hernando, J.L.: El impacto de la inteligencia artificial generativa en educación superior: Una mirada desde la ética y la integridad académica. RELIEVE: Revista Electrónica de Investigación y Evaluación Educativa 29(2), 1–19 (2023)

[18] Gasaymeh, A.-M.M., Beirat, M.A., Abu Qbeita, A.A.: University students' insights of generative artificial intelligence (AI) writing tools. Education Sciences 14(10), 1062 (2024). https://doi.org/10.3390/educsci14101062

[19] He, Q., You, X., Shen, J.: AI-generated content and academic norms: Perspectives and reflections of college students. Interactive Learning Environments, 1–15 (2025). https://doi.org/10.1080/10494820.2025.2511252

[20] Hoomanfard, M.H., Shamsi, Y.: Generative AI in dissertation writing: L2 doctoral students' self-reported use, AI-giarism, and perceived training needs. Journal of English for Academic Purposes 78, 101570 (2025). https://doi.org/10.1016/j.jeap.2025.101570

[21] Hutto, C.J., Gilbert, E.: VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media 8(1), 216–225 (2014). https://doi.org/10.1609/icwsm.v8i1.14550

[22] Jin, Y., Yan, L., Echeverria, V., Gašević, D., Martinez-Maldonado, R.: Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. Computers and Education: Artificial Intelligence 8, 100348 (2025). https://doi.org/10.1016/j.caeai.2024.100348

[23] Johnston, H., Wells, R.F., Shanks, E.M., et al.: Student perspectives on the use of generative artificial intelligence technologies in higher education. International Journal for Educational Integrity 20, 2 (2024). https://doi.org/10.1007/s40979-024-00149-4

[24] Johnson, R.B., Onwuegbuzie, A.J., Turner, L.A.: Toward a definition of mixed methods research. Journal of Mixed Methods Research 1(2), 112–133 (2007). https://doi.org/10.1177/1558689806298224

[25] Kozinets, R.V.: Netnography: The Essential Guide to Qualitative Social Media Research, 3rd edn. SAGE Publications, London (2020). https://doi.org/10.4135/9781036233846

[26] Lancaster, T., Clarke, R.: Contract cheating: The outsourcing of assessed student work. In: Bretag, T. (ed.) Handbook of Academic Integrity. Springer, Singapore (2016). https://doi.org/10.1007/978-981-287-098-8_17

[27] Li, Y., Sha, L., Yan, L., Lin, J., Raković, M., Galbraith, K., Lyons, K., Gašević, D., Chen, G.: Can large language models write reflectively? Computers and Education: Artificial Intelligence 4, 100140 (2023). https://doi.org/10.1016/j.caeai.2023.100140

[28] Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., Zou, J.: GPT detectors are biased against non-native English writers. Patterns 4(7), 100779 (2023). https://doi.org/10.1016/j.patter.2023.100779
[29] Lund, B., Mannuru, N.R., Teel, Z.A., Lee, T.H., Ortega, N.J., Simmons, S., Ward, E.: Student perceptions of AI-assisted writing and academic integrity: Ethical concerns, academic misconduct, and use of generative AI in higher education. AI Education 1(1), 2 (2025). https://doi.org/10.3390/aieduc1010002
[30] McDonald, N., Johri, A., Ali, A., Collier, A.H.: Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. Computers in Human Behavior: Artificial Humans 3, 100121 (2025). https://doi.org/10.1016/j.chbah.2025.100121
[31] McLean, N., Purton, I., Tan, W., Zhang, J.: Chinese international students' collective navigation of their study abroad experiences: Analysis of Xiaohongshu posts about UK university life. Journal of International Students 16(8), 221–240 (2026). https://doi.org/10.32674/gen6be73
[32] Mu, Y.: Research on the dissemination effect of AIGC-generated content on new media platforms: A case study of Xiaohongshu. Media Studies 1(2), 1–5 (2024)
[33] Nguyen, K.V.: The use of generative AI tools in higher education: Ethical and pedagogical principles. Journal of Academic Ethics 23, 1435–1455 (2025). https://doi.org/10.1007/s10805-025-09607-1
[34] Newton, P.M.: How common is commercial contract cheating in higher education and is it increasing? A systematic review. Frontiers in Education 3, 67 (2018). https://doi.org/10.3389/feduc.2018.00067
[35] Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T., Tacchi, J.: Digital Ethnography: Principles and Practice. SAGE Publications, London (2016)
[36] Power, M.: The Audit Society: Rituals of Verification. Oxford University Press, Oxford (1999). https://doi.org/10.1093/acprof:oso/9780198296034.001.0001
[37] Qian Gua Data: 2025 huoyue yonghu yanjiu baogao: Xiaohongshu pingtai [2025 active-user research report: Xiaohongshu platform]. https://www.qian-gua.com/information/detail/3149. Last accessed 1 Jun 2026. (in Chinese)
[38] Ren, X., Zheng, W., Zhang, M.: Navigating cognitive dissonance: Master's students' experiences with ChatGPT in dissertation writing. Frontiers in Psychology 16, 1542559 (2025). https://doi.org/10.3389/fpsyg.2025.1542559
[39] Sadasivan, V.S., Kumar, A., Balasubramanian, S., Wang, W., Feizi, S.: Can AI-generated text be reliably detected? arXiv preprint arXiv:2303.11156 (2023). https://doi.org/10.48550/arXiv.2303.11156
[40] Stieglitz, S., Dang-Xuan, L.: Emotions and information diffusion in social media: Sentiment of microblogs and sharing behavior. Journal of Management Information Systems 29(4), 217–248 (2013). https://doi.org/10.2753/MIS0742-1222290408
[41] Strathern, M.: "Improving ratings": Audit in the British University system. European Review 5(3), 305–321 (1997). https://doi.org/10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4
[42] Sui, X., Lin, Q., Wang, Q., Wan, H.: Who will benefit from AIGC: An empirical study on the intentions to use artificial intelligence generated content in higher education. Education and Information Technologies (2025). https://doi.org/10.1007/s10639-025-13590-w
[43] Tang, Y., Su, L.: Graduate education in China meets AI: Key factors for adopting AI-generated content tools. Libri 75(1), 81–96 (2025). https://doi.org/10.1515/libri-2024-0079
[44] Wankhade, M., Rao, A.C.S., Kulkarni, C.: A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review 55, 5731–5780 (2022). https://doi.org/10.1007/s10462-022-10144-1
[45] Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., Waddington, L.: Testing of detection tools for AI-generated text. International Journal for Educational Integrity 19, 26 (2023). https://doi.org/10.1007/s40979-023-00146-z
[46] Wei, G., Tinmaz, H.: A study on the impact of Xiaohongshu on university students' consumption behavior. Journal of Asian Pacific Communication 35(1), 4–28 (2025). https://doi.org/10.1075/japc.00113.wei
[47] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., et al.: Chain-of-thought prompting elicits reasoning in large language models. In: Advances in Neural Information Processing Systems, vol. 35, pp. 24824–24837 (2022)
[48] Xinhua News Agency: Lunwen AIGC lü, jiance biaozhun kexue ma? [Is the AIGC rate of papers scientifically measured?]. Quanzhou Wang, 23 May 2026. https://www.qzwb.com/gb/content/2026-05/23/content_9232318.htm. Last accessed 1 Jun 2026. (in Chinese)
[49] Yan, G., Zhu, C., Li, T., Yip, J., Ni, J.: Assessing the impact of generative AI on undergraduate thesis quality: A comparative study of students and teachers. PLOS ONE 21(4), e0347653 (2026). https://doi.org/10.1371/journal.pone.0347653
[50] Zhao, S., Chen, X., Huang, V.H., Rollins, M., Carratù, M., Shallari, I.: Students' use and attitudes toward generative artificial intelligence: A comparative study between the UK and China. In: Proceedings of the 58th Hawaii International Conference on System Sciences, pp. 4933–4940 (2025)
[51] Zhao, Y.: Universities regulate use of AI in writing. China Daily, 11 June 2024. https://www.chinadaily.com.cn/a/202406/11/WS66678d97a31082fc043cbbfb.html. Last accessed 1 Jun 2026
[52] Zhong, C.: The dilemma of academic writing in the AI era: Technological empowerment and the tension of regulation — A topic modeling study based on Xiaohongshu user discussions. In: 2025 International Workshop on Artificial Intelligence and Education (WAIE), Yokohama, Japan, pp. 388–391. IEEE (2025). https://doi.org/10.1109/WAIE67422.2025.11381059