Yuqing (Kyra) Xing; yuqixi@usc.edu; USC ID: 5407161528

What People Say About AIED on Bluesky: Themes, Tone, and Change Over Time

Note. I originally planned to analyze Reddit posts to study parents' concerns and mediation strategies related to children's TikTok use. However, Reddit did not approve my researcher API access request, so I transitioned to Bluesky, which supports programmatic data retrieval. Given Bluesky's active technology- and education-oriented communities, I reframed the study to focus on AIED discourse.

Project Description
Public discourse on AI in education (AIED) evolves quickly, often blending practical discussions of classroom use with broader debates about policy, assessment, and academic integrity. This project analyzes publicly available Bluesky posts related to AIED with four objectives: (1) to construct an analysis-ready corpus by cleaning and standardizing post text, including systematic removal of duplicates, spam, and low-information content; (2) to identify prominent themes using complementary methods, including frequency-based extraction of salient terms/phrases and unsupervised topic modeling; (3) to estimate affective orientation using a transparent baseline sentiment approach (VADER); and (4) to examine temporal dynamics by tracking shifts in both topic prevalence and sentiment over time.

Three research questions guide the analysis:
- RQ1: What are the dominant themes in AIED posts, and how prevalent is each theme in the corpus?
- RQ2: How does sentiment vary across rule-based education-context tags (e.g., curriculum, assessment, policy, edtech tools)?
- RQ3: How do topic prevalence and sentiment shift over time in AIED-related posts?

Data Collection and Cleaning
The dataset was collected from Bluesky, a public microblog-style social media platform that provides a timely venue for observing emerging public discourse, including conversations among education- and technology-oriented communities about generative AI (GenAI) and its implications for learning and teaching. We used the platform's official AT Protocol XRPC endpoints via HTTPS API calls to retrieve data. Posts were queried through *app.bsky.feed.searchPosts* (on *https://bsky.social*), with authentication handled through *com.atproto.server.createSession* using a Bluesky handle and app password stored in environment variables (from a local *.env* file). Posts were selected using a query strategy that enumerated the cross product of an AI-related term set and an education-related term set. The collection window spanned from 2024-01-01 to 2025-12-10.

The data cleaning process was designed to minimize duplicates, advertisements, and low-information posts. First, the pipeline ingests the newline-delimited JSON file and removes duplicate records based on a unique post identifier (*id* or *uri*), ensuring that the same post is not counted multiple times due to repeated retrieval across overlapping queries. Second, posts are filtered to English using automated language detection (langdetect), which helps maintain linguistic consistency for topic modeling and sentiment analysis. Third, a minimum-length threshold (> 50 chars) is applied to reduce noise from very short posts. Fourth, the pipeline performs text-level de-duplication to eliminate repeated messages that differ only superficially. Post text is normalized by lowercasing, removing URLs, user mentions, emoji characters, and collapsing whitespace; posts that share the same normalized text are treated as duplicates and only one instance is retained. Fifth, potential advertisements and platform spam are removed using a conservative heuristic that flags posts containing promotional or client-solicitation language (e.g., "open to clients," "copywriter"), explicit advertising hashtags (e.g., #ad, #sponsored), heavy link patterns, and recurring spam templates (e.g., "Bluesky's Top 10 Trending Words").

Data Analysis

To address RQ1 (primary topics/themes and their prevalence), we combined a rule-based classification with corpus-driven text mining and unsupervised topic modeling. First, we quantified education-context prevalence using dictionary-based context tags spanning policy, assessment, curriculum, support services, edtech tools, infrastructure, professional development, and stakeholders. This approach serves as a transparent first-pass taxonomy that maps discourse into widely recognized domains in AIED research and practice. For example, the *policy* tag captures governance and compliance language (e.g., guidelines, ethics, FERPA), *assessment* captures evaluation and integrity concerns (e.g., exams, grading, cheating/plagiarism), and *curriculum* captures instructional design and coursework artifacts (e.g., syllabi, lesson plans, assignments). Second, to surface themes not fully captured by predefined tags, we conducted a top terms/phrases analysis (TF–IDF). Third, we applied Non-negative Matrix Factorization (NMF) to identify latent themes from the document–term matrix. Topic prevalence was operationalized by assigning each post to its most dominant topic (i.e., the topic with the maximum NMF weight), enabling straightforward prevalence estimates and temporal aggregation.

To address RQ2 (sentiment differences across topics and context tags), we computed sentiment scores using Valence Aware Dictionary and Sentiment Reasoner (VADER), a rule-based method developed for short, informal texts (e.g., social media). VADER produces a compound polarity score on $[-1,1][-1,1][-1,1]$, where values closer to +1 reflect more positive affect (e.g., endorsement, optimism), values closer to -1 reflect more negative affect (e.g., criticism, frustration), and values near 0 reflect neutral or mixed tone. In this project, the compound score is used as an interpretable proxy for affective tone rather than a definitive measure of stance, because AIED posts may contain complex rhetorical structures (e.g., sarcasm, conditional endorsements) that are difficult to capture with rule-based methods. Nevertheless, VADER provides a consistent and transparent baseline for comparisons across contexts and topics.

To address RQ3 (temporal evolution of topic prevalence and sentiment), we leveraged post timestamps to compute time-indexed aggregates. We parsed *created_at* into a consistent UTC datetime format and computed daily posting volume to contextualize fluctuations in overall activity. We then calculated the daily share of each dictionary-based context tag and visualized changes in bucket composition over time, allowing us to observe whether particular educational contexts become more salient during specific periods. Using the same aggregation strategy, we tracked topic prevalence trajectories based on NMF topic assignments, which reveals how latent themes expand or contract over time. Finally, we computed daily (or weekly) average sentiment and examined its evolution alongside changes in topical composition.

Data Visualization and Results

For RQ1, we first report education-context bucket prevalence as a transparent, rule-based overview of where AIED discourse is situated within educational practice and institutions. The prevalence bar chart (Figure 1) indicates that AIED discussion is concentrated in instructional and evaluative domains: *curriculum* is the most prevalent bucket (14%), followed by *assessment* (11%). All other buckets occur at substantially lower rates. We then examine context co-occurrence using a bucket-by-bucket heatmap (Figure 2). Most posts are tagged with a single bucket. Among off-diagonal overlaps, the strongest co-occurrence is between curriculum and assessment, suggesting that instructional planning and classroom practice are frequently discussed in direct relation to grading, evaluation, and integrity concerns.

We next examined salient language patterns using TF–IDF with unigrams and bigrams. At the corpus level, the highest-scoring terms and phrases reflect recurring ways users discuss GenAI in educational settings, including references to specific systems (e.g., ChatGPT), institutional settings and actors (e.g., classroom, teacher, students), and common academic activities (e.g., writing, assignments, grading). We also computed TF–IDF within context buckets to compare vocabulary across educational domains. Consistent with expectations, *assessment*-tagged posts surface phrasing associated with evaluation and

integrity (e.g., grading- and cheating-related language), whereas *curriculum*-tagged posts emphasize planning and instructional design language; *policy*-tagged content tends to include governance and guideline-oriented wording.

For RQ2, we analyzed differences in affective tone across both context tags and latent topics using VADER's compound polarity score. We first visualized the overall sentiment distribution with a histogram (Figure 3) to establish the baseline: most posts cluster near neutral, with thinner positive and negative tails. We then compared sentiment across education-context buckets using boxplots (Figure 4). Across buckets, central tendency remains neutral to mildly positive with substantial dispersion. While some buckets show slightly higher median sentiment (e.g., policy, support services, infrastructure, professional development) and others trend modestly lower (e.g., assessment, stakeholders), these differences are small relative to the wide within-bucket variability. The presence of long tails across all buckets indicates that strongly positive and strongly negative posts appear in every context, underscoring that AIED discourse contains both enthusiasm and critique regardless of institutional framing.

For RQ3, temporal analyses indicate that AIED discourse on Bluesky is dynamic and appears event-responsive. The daily posting volume series (Figure 5) shows relatively low activity through much of 2024, followed by a clear upward shift beginning in late 2024 and continuing through 2025, punctuated by multiple sharp spikes. The bucket composition plot over time (Figure 6) further indicates that curriculum and assessment remain consistently represented, but their daily shares fluctuate and occasionally spike, suggesting short intervals during which specific contexts become disproportionately salient. Together, these results imply that even when a stable set of broad themes dominates, the relative emphasis across contexts varies over time in ways consistent with attention cycles and episodic triggers.

Discussion and Future Work
Findings from the thematic analyses (RQ1) suggest that AIED discourse on Bluesky is dominated by practical, classroom-facing concerns, especially instructional design and assessment-related issues. This pattern aligns with recent syntheses and policy guidance emphasizing that GenAI adoption becomes most visible (and most contested) at instructional "pressure points," such as designing assignments, supporting student writing, and evaluating student work under emerging integrity constraints (Bond et al., 2024; Miao & Holmes, 2023; Cotton et al., 2024). The context co-occurrence evidence further reinforces this interpretation: *curriculum* and *assessment* are not only the most prevalent buckets but also the most salient intersection, suggesting that instructional decisions are frequently discussed alongside grading, evaluation, and integrity considerations (Cotton et al., 2024; Eaton, 2023). TF–IDF and topic modeling results sharpen this picture by showing that high-salience language is anchored in educational actors/settings (e.g., teachers, students, classroom) and GenAI references (e.g., AI, ChatGPT).

Findings from the sentiment analysis (RQ2) indicate that broad education-context tags do not sharply partition affective tone: policy, curriculum, assessment, and tools can all contain supportive, critical, and mixed posts. This is consistent with literature noting that GenAI in education is debated through competing value frames, innovation, access, and productivity on the one hand, and risks related to integrity, equity, privacy, and governance on the other (Bittle & El-Gayar, 2025; Miao & Holmes, 2023; Eaton, 2023). The dispersion observed across buckets suggests that sentiment is often driven more by the post's stance and framing (e.g., "helpful support" versus "integrity threat") than by the category label itself. Substantively, Bluesky AIED discourse reflects ongoing norm negotiation about what constitutes legitimate AI assistance in learning and assessment.

Temporal analyses (RQ3) show that AIED posting volume and composition are dynamic and plausibly event-responsive: activity increases substantially from late 2024 into 2025, with spikes that resemble attention-cycle behavior documented in classic public-issue dynamics. Even when *curriculum* and *assessment* remain consistently prominent, their daily shares fluctuate, suggesting that the conversation's

emphasis shifts over time, potentially tracking school-year rhythms, news cycles, policy announcements, or platform-level dynamics (Miao & Holmes, 2023).

Future work can strengthen both methodological rigor and substantive insight. Methodologically, the rule-based context tags and ad/spam filters should be evaluated against a manually annotated subset to quantify precision/recall and iteratively refine keyword dictionaries. Sentiment analysis could be extended beyond a transparent baseline (VADER) to transformer-based methods fine-tuned for stance or education-specific sentiment, which may better capture irony, mixed valence, and policy-oriented evaluative language. Topic modeling could be augmented by testing alternative approaches such as embedding-based models (e.g., BERTopic) and systematically evaluating topic coherence and stability across preprocessing and random seeds.
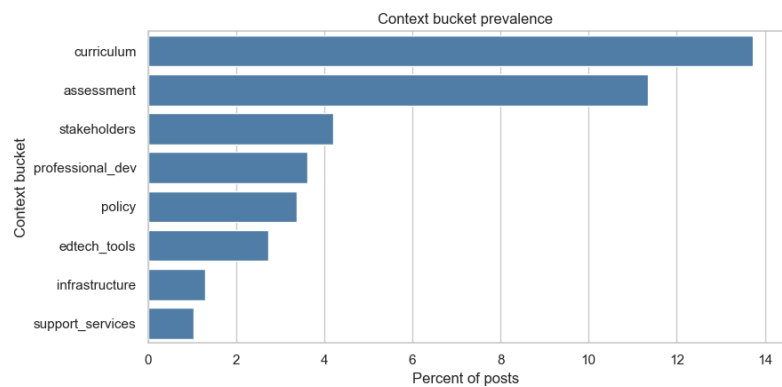
Figure 1
Context Bucket Prevalence
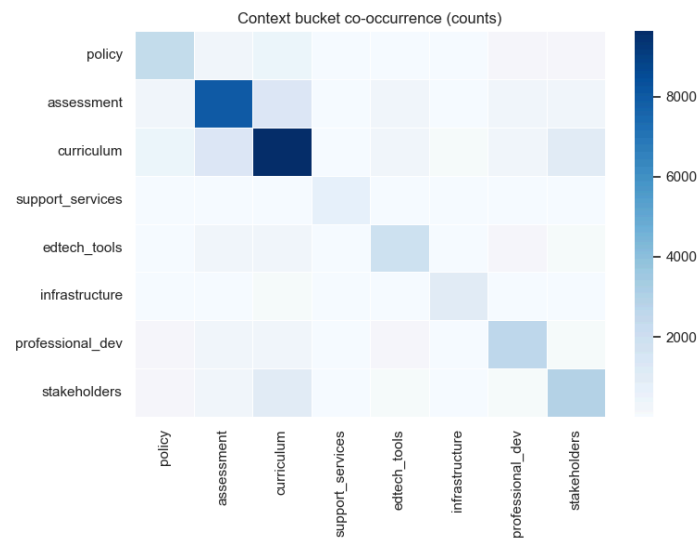

Figure 2
Context Bucket Co-occurrence Heatmap
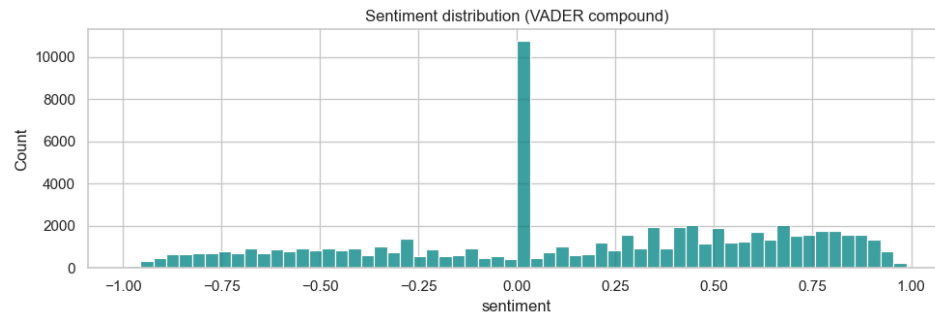

Figure 2
Overall Sentiment Distribution

Sentiment distribution (VADER compound)

Figure 3

Sentiment Analysis by Context Buckets
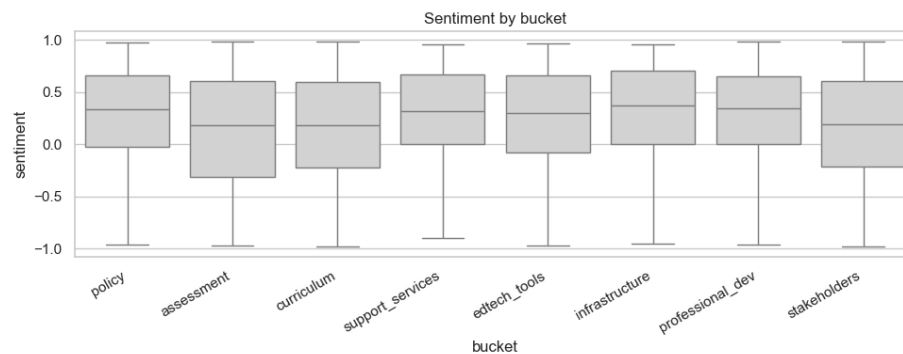

Sentiment by bucket

Figure 4

Temporal Analysis of Daily Posting Volume
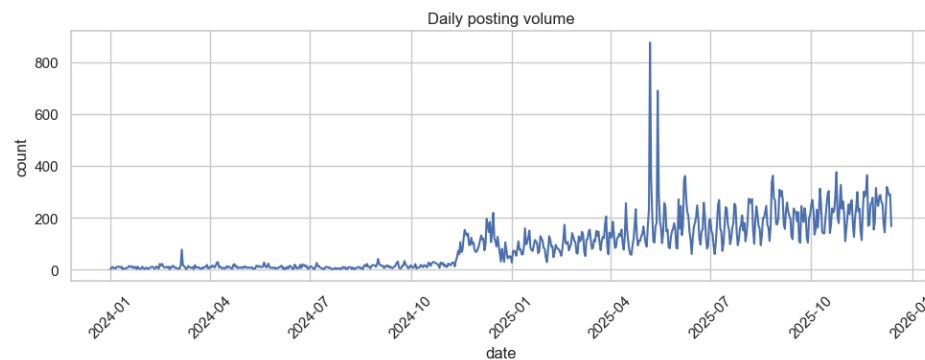

Daily posting volume

Reference

Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: a systematic review and research agenda. *Information, 16*(4), 296.

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*, 4.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239.

Eaton, S. E. (2023). Postplagiarism: transdisciplinary ethics and integrity in the age of artificial intelligence and neurotechnology. *International Journal for Educational Integrity, 19*, 23.

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO.