Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Divergent research orientations of affective and emotion concepts in artificial intelligence education through a bibliometric analysis

Hsiang Ni Tsai
National Tsing Hua University
Sheng Yi Wu
digschool@gmail.com

National Tsing Hua University

Research Article

Keywords: affective computing, emotion recognition, artificial intelligence in education, bibliometric analysis, generative AI, keyword co-occurrence, content analysis, educational technology

Posted Date: June 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9878709/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Discover Artificial Intelligence on November 2nd, 2026. See the published version at https://doi.org/10.1007/s44163-026-02642-w.

Divergent research orientations of affective and emotion concepts in artificial intelligence education through a bibliometric analysis

Hsiang-Ni,Tsai Sheng-Yi,Wu*
National Tsing Hua University, Department of Education and Learning Technolog, Hsinchu, Taiwan
National Tsing Hua University, Center for Teacher Education , Hsinchu, Taiwan
iamwiifit1@gmail.com
digschool@gmail.com*

Abstract

This study employs bibliometric analysis to systematically review research trends in AI-driven affective and emotional education published in the Scopus database from 2020 to 2025. Following the PRISMA 2020 guidelines, a dual-track keyword search strategy was adopted, ultimately yielding 401 documents for analysis. A five-dimensional coding framework was established encompassing research purpose, measurement channel, research participants, learning environment, and subject domain, with coding reliability confirmed through a triangulation procedure.

Three principal findings emerged. First, keyword co-occurrence analysis revealed that the field has coalesced into four core clusters — technical recognition, psychological experimentation, specialized applications, and instructional practice — signaling that AI affective research has matured into a structurally coherent knowledge ecosystem. Second, longitudinal analysis identified 2023 as a critical inflection point: annual publication volume grew from 26 articles in 2022 to 197 in 2025, with the rise of generative AI driving a research shift from passive sensing and detection toward conversational affective scaffolding. Third, chi-square tests across all five research dimensions yielded statistically significant differences, confirming that "Affective" research is oriented toward educational practice while "Emotion" research is oriented toward technological development.

This study proposes the Dual Knowledge Production Model of Affective and Emotion AI Research, clarifying the systematic academic division of labor between the two dominant terms and providing an empirically grounded framework to facilitate cross-disciplinary collaboration.

Keywords: affective computing, emotion recognition, artificial intelligence in education, bibliometric analysis, generative AI, keyword co-occurrence, content analysis, educational technology

1. Background and Motivation

With the rapid advancement of artificial intelligence and machine learning in recent years, considerable research has focused on the integration of AI and machine learning in adaptive learning — that is, providing personalized learning experiences tailored to students' needs, preferences, and progress (Gligorea, 2023). Prior literature has also documented a shift in AI applications in education from behaviorist paradigms that follow prescribed learning pathways and emphasize cognitive exploration, toward more human-centered approaches that incorporate the affective dimension (Ouyang & Jiao, 2021). As Bloom et al. (1956) emphasized, the affective domain is an indispensable component of holistic learning. Moreover, with the normalization of digital learning environments in the post-pandemic era, learners' mental health and emotional regulation have emerged as critical factors influencing learning outcomes (Rao & Rao, 2021). To address this research gap, the present study aims to employ bibliometric analysis to systematically examine the research dynamics of AI applications in the affective domain of education.

This field did not develop gradually into its current form; rather, it underwent a dramatic technological leap with the explosion of generative AI in 2023. Prior to 2022, research largely focused on using sensors for emotion recognition and constructing higher-dimensional affective models (Zhang et al., 2020). However, with the advent of large language models such as ChatGPT, the role of AI rapidly transformed into that of an interactive partner capable of natural dialogue and emotional support (Bansal et al., 2024). This technological shift from unidirectional detection to bidirectional interaction has fundamentally reshaped human–computer interaction (Kasneci et al., 2023). Nevertheless, existing reviews either focus on a single term (Wei & Li, 2024) or fail to cover recent developments following the rise of GenAI after 2023 (Jiang & Huang, 2023), leaving a gap in systematic reviews that simultaneously integrate dual-track terminological comparison and recent trend evolution.

Furthermore, given the inherently interdisciplinary nature of AI educational applications, engineering and educational-psychological perspectives have increasingly converged in recent years, blurring the boundary between "affective" and "emotion." Terminological conflation remains prevalent in the literature (Wei & Li, 2024; Pei et al., 2024), with no clear consensus on the specific differences between these two key terms in terms of application domains, technical levels, or research targets. It can be inferred that engineering-oriented system developers, lacking contextual awareness of actual classroom settings, may tend to optimize systems based primarily on instantaneous emotion recognition rates, overlooking the long-term and complex nature of affective development in authentic educational contexts. Conversely,

education-oriented researchers with insufficient understanding of the technical limitations of underlying algorithms may articulate affective support expectations that are difficult to realize within existing computational frameworks. This study therefore attempts to clarify the systematic differences between "Affective" and "Emotion" in AI education research. Rather than discussing terminological definitions at a purely theoretical level, this study establishes a five-dimensional coding framework encompassing research purpose, measurement channel, research participants, learning environment, and subject domain. Quantitative content analysis and statistical testing are applied to both keyword groups to objectively document differences in research orientations and to assist the academic community in building a cross-disciplinary collaborative framework.

In summary, this study seeks to answer three core research questions:

RQ1. What are the core research theme clusters in the field of "AI in affective/emotional education" in the Scopus database during 2020–2025?

RQ2. How have research trends in this area evolved over time? Specifically, how has the emergence of generative AI manifested in shifts and expansions of research hotspots?

RQ3. Do documents retrieved using "Emotion" versus "Affective" as keywords exhibit significant differences in application domains, technical characteristics, or research target focus?

2. Literature Review

Research on affective and emotional dimensions in the AI education field is rooted in three theoretical traditions — educational science, engineering, and psychology — each of which has developed distinct research orientations and methodological approaches. At the educational level, Bloom et al. (1956) established the affective domain as one of the three core domains of their taxonomy of educational objectives, encompassing interests, attitudes, values, and aesthetic appreciation, thereby laying the curricular foundation for affective education. At the engineering level, Picard's (2000) concept of "affective computing" endowed computers with the capacity to recognize and express emotions, pioneering a technical pathway for machines to understand human affect. At the psychological level, Pekrun's (2006) Control-Value Theory of Achievement Emotions positioned emotions as multi-component coordinated systems encompassing affective, cognitive, motivational, expressive, and physiological processes, providing a theoretical framework for measuring emotions in educational contexts. More recent research has further indicated that AI systems have shifted from mere emotion

recognition toward regulating students' sense of control and value perception through real-time feedback (Yuvaraj et al., 2025), signaling an accelerating integration of the three traditions. However, precisely because this field spans three academic communities, research topics are highly dispersed, making it difficult to grasp the full picture from individual studies. Accordingly, bibliometric cluster analysis has been recognized as an effective tool for systematically organizing large-scale interdisciplinary knowledge structures (Arruda et al., 2022), and the present study adopts this method to address RQ1, presenting the current state of the knowledge landscape through objective data.

Building upon these three theoretical traditions, rapid AI development has further expanded affective education practice and witnessed a dramatic change in recent years. Prior to 2022, research predominantly focused on passive emotion recognition using multimodal sensors, with AI playing the role of observer and diagnostician (Wang et al., 2022; Wei & Li, 2024). However, with the rise of large language models such as ChatGPT, AI's role rapidly transformed into that of an interactive partner capable of natural dialogue and emotional support (Bansal et al., 2024; Kasneci et al., 2023). Empirical research has further confirmed that large language models perform comparably to humans in emotional awareness (Elyoseph et al., 2023; Sharma et al., 2023). This technological shift from unidirectional signal detection to bidirectional affective interaction (Zhang et al., 2024; Wang et al., 2026) marks a fundamental transformation in AI's role in educational emotional support. Meanwhile, this transformation has also introduced new research questions: as AI becomes an affective interaction partner for learners, the risks of human–AI emotional dependency and the ethical boundaries of such relationships also require clarification (Fang et al., 2025). Yet existing bibliometric studies tend to adopt analytical intervals spanning a decade or more (Yuvaraj et al., 2025; Wei & Li, 2024), making it difficult to sensitively capture the structural reorganization of research hotspots triggered by the rise of GenAI after 2023. This study therefore focuses on the recent interval from 2020 to 2025 to precisely highlight the impact of this technological transition on research hotspots, addressing RQ2.

With regard to keyword delineation, existing literature indicates that "Affective" and "Emotion" frequently exhibit high conceptual overlap and are used interchangeably (Dilekçi & Manap, 2022; Jiang & Huang, 2023; Yuvaraj et al., 2025). In terms of academic definitions, "Affective" originates from the intersection of educational science and engineering: in Bloom et al.'s (1956) taxonomy, it denotes the broad affective domain encompassing attitudes, values, and interests, while in Picard's (2000) affective computing framework, it refers to the system-

level capacity for affective state modeling. "Emotion," by contrast, is more closely aligned with the psychological tradition — as in Ekman's (1992) basic emotion theory and Pekrun's (2006) achievement emotion theory — focusing on specific, detectable psychological states and processes. The two terms differ at their academic origins, corresponding to two independent knowledge pathways: the educational tradition originating with Bloom et al. (1956) and the engineering-psychological tradition originating with Ekman (1992) and Picard (2000). This study contends that this represents not merely terminological conflation, but rather the natural manifestation in the literature of a fundamental difference in the knowledge-production logics of two academic communities. It is therefore expected that these differences will leave statistically verifiable systematic traces across research purpose, measurement approach, research targets, and application domains, addressing RQ3.

Surveying existing literature, three gaps remain to be filled: existing bibliometric studies primarily conduct analyses using a single database or single time-point, lacking systematic integration of core theme clusters in AI affective education; existing studies predominantly adopt analytical intervals of ten years or more, making it difficult to sensitively capture the specific impact of GenAI's rise after 2023 on research hotspot structures; and no study has yet empirically tested the systematic differences in research orientations between the two key terms "Affective" and "Emotion" using statistical methods, with most existing bibliometric research focusing on a single term and lacking a dual-track comparative perspective. This study aims to simultaneously address all three gaps, corresponding respectively to RQ1 through RQ3.

3. Methodology

3.1 Data Source and Search Strategy

This study uses the Scopus database, which maintains rigorous journal inclusion standards ensuring academic quality, and provides comprehensive indexing and bibliographic data suitable for large-scale science mapping analysis (Arruda et al., 2022). To precisely identify relevant literature, a three-layer Boolean logic search string was employed. The specific search strings are as follows:

Affective search string:

```
TITLE-ABS-KEY( education ) AND TITLE-ABS-KEY( ai ) AND TITLE-ABS-KEY( affective )AND PUBYEAR > 2019 AND PUBYEAR < 2026 AND DOCTYPE( ar ) AND LANGUAGE( english )
```

Emotion search string:

```
TITLE-ABS-KEY( education ) AND TITLE-ABS-KEY( ai ) AND TITLE-ABS-KEY( emotion )AND PUBYEAR > 2019 AND PUBYEAR < 2026 AND DOCTYPE( ar ) AND LANGUAGE( english )
```

The search scope was limited to article titles, abstracts, and keyword fields. Regarding the selection of affective/emotion keywords, this study included "Affective" and "Emotion" while explicitly excluding "Sentiment" for the following reasons:

First, "Affective" has dual theoretical roots: in Bloom et al.'s (1956) taxonomy of educational objectives, the Affective Domain is one of the three core learning domains, encompassing attitudes, values, and interests; in Picard's (2000) affective computing framework, it refers to the system-level capacity to model and respond to emotional states. Both refer to deep, sustained affective processes that are highly relevant to educational contexts.

Second, "Emotion" has a strong empirical foundation in psychology, as Ekman's (1992) basic emotion theory focuses on specific, identifiable psychological states, and Pekrun's (2006) achievement emotion theory addresses emotional regulation processes in learning contexts, both aligning with research needs centered on student psychological wellbeing in educational settings.

Third, "Sentiment" is primarily applied in natural language processing for text polarity analysis, focusing on detecting the positive, negative, or neutral orientation of textual information. This constitutes a surface-level semantic computational task that does not involve affective development or emotional regulation in the educational-psychological sense and was therefore excluded.

3.2 Data Collection Procedure

This study followed PRISMA 2020 standards for data collection, employing a dual-track parallel strategy by separating Scopus search results into independent "Affective" and "Emotion" keyword groups to address RQ3. Literature searches were conducted on November 28, 2025.

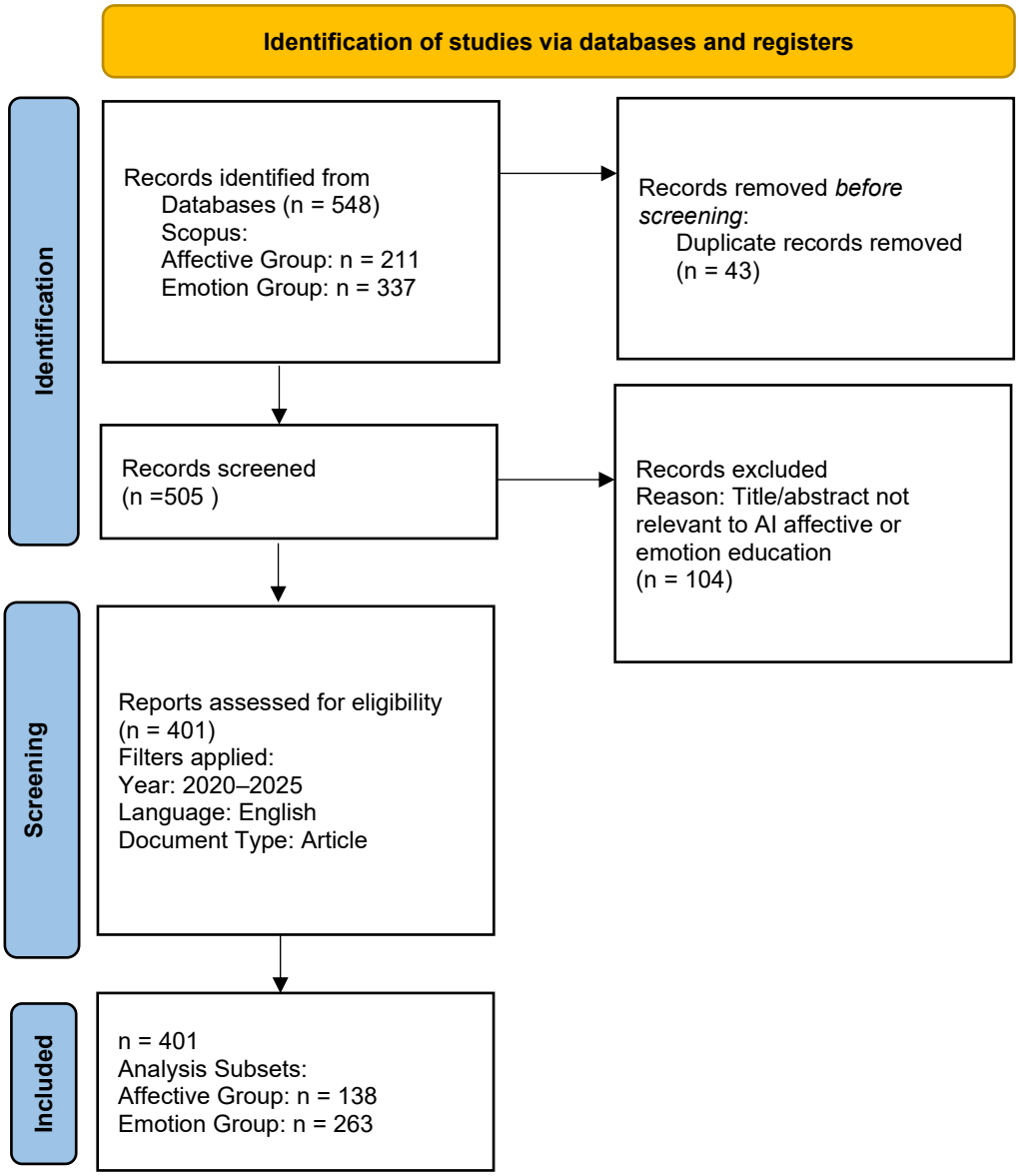
The search scope was limited to the title, abstract, and keyword fields (TITLE-ABS-KEY). The time range was set from 2020 to 2025 to capture research changes before and after the rise of generative AI; screening criteria were restricted to English-language journal articles (Language: English; Document Type: Article) to ensure the academic quality and homogeneity of the analytical sample.

The complete data collection procedure followed PRISMA 2020 specifications, as

shown in Figure 3-1. Initial search results for both groups totaled 548 documents (Affective group: 211; Emotion group: 337). After removing 43 duplicate documents, 505 documents entered the screening procedure. The screening stage involved manual review based on the relevance of titles and abstracts, with 104 documents excluded as irrelevant to AI affective or emotional education applications. A final total of 401 documents were included in the analysis (Affective group: 138; Emotion group: 263).

Figure 3-1.

PRISMA 2020 Flow Diagram



3.3 Data Analysis and Visualization

This study employed VOSviewer software for keyword co-occurrence network analysis.

VOSviewer is designed specifically for constructing visualized bibliometric networks, excels at processing large bibliographic datasets, and presents complex knowledge structures through cluster maps and overlay visualizations (Arruda et al., 2022).

3.3.1 Overall Knowledge Map and Longitudinal Analysis

To faithfully represent the diverse developmental landscape of AI affective education, this study preserved original terminology without enforcing synonym merging within the software when constructing the overall keyword co-occurrence network. Given that generative AI terms (e.g., ChatGPT, GenAI, LLM) represent emerging and rapidly evolving topics whose terminology has not yet fully standardized, excessive merging could obscure micro-trends in emerging vocabulary. The minimum keyword occurrence threshold was therefore set at 3 to capture marginally emerging but important nodes, fully preserving the richness of the bibliographic network. Longitudinal analysis introduced a temporal dimension into the network through VOSviewer's overlay visualization function, using node color gradients to represent the average publication year of each keyword.

3.3.2 Top Ten Keyword Comparison for Dual Groups

For comparing the core keyword focus of the "Affective" and "Emotion" groups, this study adopted a post-hoc weighted merging approach. Specifically, raw VOSviewer analysis data for both groups were first exported separately, then synonym-weighted merging calculations were performed externally, summing occurrence frequencies of terms denoting the same concept — for example, merging "AI," "Artificial Intelligence," and "Computational Intelligence"; and merging "GenAI," "ChatGPT," and "Generative AI." The top ten high-frequency keywords for each group were then ranked by merged weighted frequencies, serving as the first-level comparative evidence for differences in research focus between the two terms.

3.4 Five-Dimensional Content Analysis

3.4.1 Five-Dimensional Content Analysis Framework

To systematically compare the differences in research orientations between the "Affective" and "Emotion" groups, this study established a content analysis coding framework covering five dimensions, with the following classification categories:

(1) Research Purpose (P): five categories — recognition/diagnosis (P1), affective tutoring/adaptation (P2), empathic partner/GenAI (P3), evaluation of existing tool impact (P4), and unspecified/purely theoretical (P5).

(2) Measurement Channel (M): six categories — traditional questionnaire/self-report (M1), physiological sensors (M2), facial expression/speech (M3), text sentiment analysis (M4), multimodal fusion (M5), and unspecified/purely algorithmic (M6).

(3) Research Participants (S): five categories — K-12 (S1), higher education (S2), special education (S3), teachers/educators (S4), and unspecified (S5).

(4) Learning Environment (E): four categories — authentic classroom (E1), online/distance learning (E2), experimental/dataset (E3), and unspecified (E4).

(5) Subject Domain (L): four categories — STEM (L1), language learning (L2), single specific subject (L3), and general/non-specific subject (L4).

3.4.2 Coding Procedure

To ensure the objectivity and reliability of coding results, this study adopted a multi-rater triangulation procedure for content analysis. Specifically, the title, abstract, and keywords of each document were coded independently by two raters according to the operational definitions above, with raters blinded to each other's results to ensure independence of judgment. After independent coding was completed, the results were systematically compared: where both raters agreed, the result was directly adopted; where discrepancies arose, a third-party researcher reviewed the coding according to the operational definitions of the coding framework and adjudicated the final coding. This procedure follows the inter-rater reliability confirmation mechanism recommended by Campbell et al. (2013), aiming to reduce systematic bias from a single judgment source through cross-validation from multiple perspectives.

3.4.3 Coding Reliability

To evaluate the consistency of the triangulation procedure, this study calculated Cohen's Kappa coefficients for the initial coding of the two independent raters. As shown in Table 3-1, Kappa values ranged from .405 to .821 for the Affective group and from .160 to .684 for the Emotion group. The research participants and subject domain dimensions reached acceptable to high consistency levels, while the research purpose, measurement channel, and learning environment dimensions reached fair to acceptable levels, with the overall reliability meeting acceptable standards (Landis & Koch, 1977). The relatively low Kappa value for the research purpose dimension in the Emotion group (κ = .160) reflects higher inherent conceptual ambiguity in defining research purpose for this term.

Table 3-1.

Cohen's Kappa Reliability for Dual Groups Across Five Dimensions

Dimension	Affective	Emotion
Research Purpose	.478	.160
Measurement Channel	.405	.370
Research Participants	.821	.684
Learning Environment	.487	.313
Subject Domain	.804	.522

3.4.4 Statistical Analysis

After coding was completed, chi-square tests were used to compare the distributional differences between the two groups across the five dimensions, and Cramer's V was calculated as an effect size measure. Cohen's (1988) criteria were applied to assess the practical significance of differences, thereby testing whether the divergence in research orientations between the two keyword groups was statistically significant.

4. Results and Discussion

4.1 Comparative Analysis of Core Theme Clusters in AI Affective and Emotional Research

This study conducted keyword co-occurrence network analysis on 401 documents from the Scopus database using VOSviewer. Results showed that the field has formed four core theme clusters, each representing a distinct facet of the ecosystem spanning technical recognition, psychological experimentation, specialized applications, and instructional practice, as shown in Figure 4-1.

Green Cluster: Affective Computing and Multimodal Recognition

This cluster represents the technical foundation of the field, with core keywords including Deep Learning, Emotion Recognition, Face Extraction, and Sentiment Analysis. This indicates that research in the green cluster focuses on using algorithms to process data from different modalities — processing visual signals through Face Extraction to read facial expressions, and processing textual data through Sentiment Analysis to analyze semantic polarity. The research objective is to optimize deep learning models and improve the accuracy of systems in recognizing human facial expressions and internal emotions. However, such research is predominantly validated on laboratory datasets, remaining considerably distant from authentic classroom teaching contexts, revealing a persistent gap between technical development and educational application.

Red Cluster: Instructional Practice and Literacy Applications

This cluster revolves around AI Literacy, Pedagogy, Teaching, and Curriculum. It indicates that AI affective technologies have crossed laboratory boundaries and entered authentic educational settings. Research emphasis has shifted from pure technical validation toward exploring how AI can be integrated into curriculum design, how teachers can leverage AI to support instruction, and how to cultivate students' affective and digital literacy in AI environments. Notably, the appearance of "AI Literacy" as a keyword marks the field's transition from using AI to teach toward teaching students to understand AI, indicating that researchers have begun to recognize the importance of learners' reflective capacity regarding AI emotional mechanisms. This cluster carries the substance and foundation of curricular practice.

Dark Blue Cluster: Psychological Constructs and Human-Factors Experiments

This cluster centers on Anxiety, Questionnaire, Students, and Human Experiment. This segment focuses on exploring learners' internal psychological states and responses to human–computer interaction. Unlike purely survey-based approaches, the presence of Human Experiment indicates that researchers have begun adopting rigorous experimental designs to examine the specific effects of AI intervention on student psychological variables (such as anxiety) in controlled environments, emphasizing human-centered empirical measurement. This cluster's research orientation aligns closely with the educational psychology tradition, providing an empirical basis for causal inference regarding the effectiveness of AI interventions.

Yellow Cluster: Specialized Domains and Predictive Assessment

This cluster includes Controlled Study, Prediction, and Innovation, as well as specific application contexts such as Autism and Fear. This highlights specific applications of the field in special education and emotional warning systems. Unlike purely technical development, this type of research emphasizes the use of rigorous controlled experimental designs to examine the predictive capability and intervention effectiveness of models for specific psychological states (such as fear) or specific populations (such as individuals with autism). This cluster reflects an increasing emphasis in AI affective research on scientifically grounded empirical data, while also extending into the special education domain.

Synthesizing the knowledge structure across the four clusters, the field has developed a

well-differentiated research ecosystem: the green and dark blue clusters form the underlying technical and psychological measurement foundation, while the yellow and red clusters represent application layers extending the foundational technologies to special education settings and general instructional practice respectively. Comparative details of core themes, representative keywords, and research focus for each cluster are shown in Table 4-1.

Figure 4-1.

Keyword Co-occurrence Network Map from Scopus Database

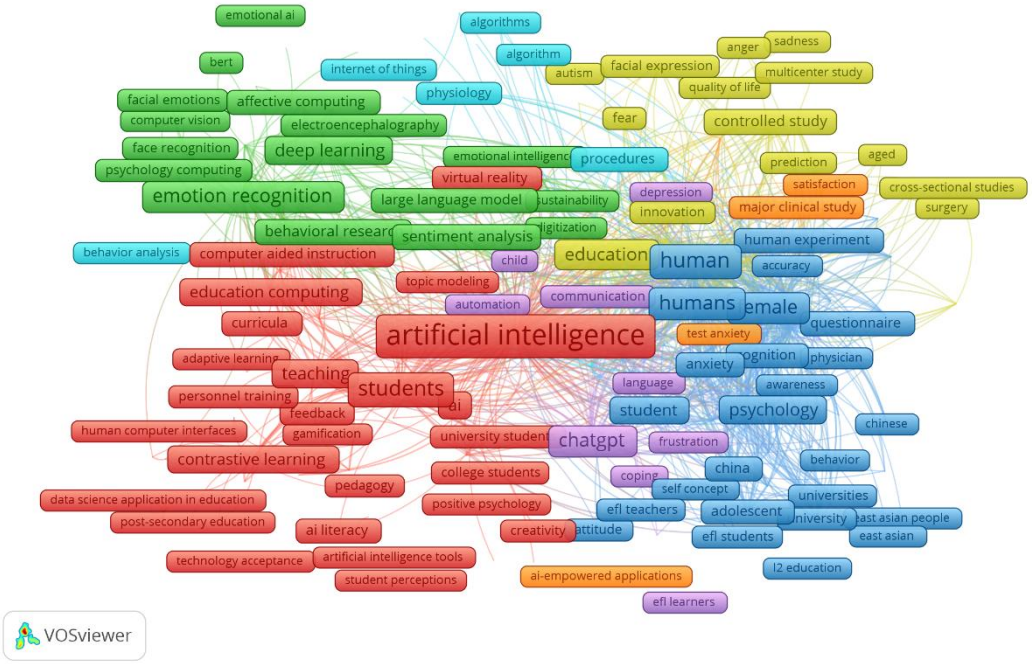


Table 4-1.

Comparative Table of Core Cluster Patterns in AI Affective and Emotional Research

Cluster	Core Theme	Representative Keywords	Research Focus and Implications
Green	Affective Computing and Multimodal Recognition	Deep Learning, Emotion Recognition, Face Recognition, Sentiment Analysis	Technical: Focused on algorithm optimization; uses multimodal data to improve machine accuracy in reading emotions.
Red	Instructional Practice and Literacy Applications	AI Literacy, Pedagogy, Teaching, Curricula	Pedagogical: Focused on classroom settings; explores AI integration into curriculum design and AI literacy development for teachers and students.
Dark Blue	Psychological Constructs and	Anxiety, Questionnaire,	Participant-centered: Focused on human psychological responses; measures

	Human-Factors Experiments	Students, Human Experiment	student anxiety and psychological states in AI environments through experiments and questionnaires.
Yellow	Specialized Domains and Predictive Assessment	Controlled Study, Prediction, Innovation, Autism, Fear	Application: Focused on special needs and effectiveness validation; develops predictive models for autism or fear emotions.

4.2 Research Trend Evolution

This study introduced a temporal dimension into the keyword co-occurrence network through VOSviewer's overlay visualization function (as shown in Figure 4-2), using node color gradients to represent the average publication year of each keyword, thereby tracking the evolution of research hotspots.

In terms of annual document distribution, the development of this field shows a pronounced exponential growth trajectory, as shown in Table 4-2. From 2020 to 2022, there were only 51 articles in total (12.7%); the count climbed sharply to 48 in 2023 (12.0%), reached 105 in 2024 (26.2%), and surged to 197 in 2025 alone (49.1%), approximately 15 times the 2020 figure. This growth trajectory closely coincides with the official release of ChatGPT at the end of 2022, indicating that the rise of generative AI is the core driver behind the dramatic expansion of research output in this field.

Further analysis of node color distribution in the longitudinal map allows the research evolution to be summarized into the following four stages:

4.2.1 Conceptual Emergence (2020–2022)

The early research period was dominated by cool-toned nodes, with literature primarily focused on Emotional AI and Data Analysis. Research at this stage was in the conceptual definition and exploratory phase, with the core direction being the transformation of abstract emotional concepts into computable data representations, laying the foundation for subsequent technical development. This stage accounted for only 51 articles, representing 12.7% of the total sample, indicating that the field remained a niche interdisciplinary topic prior to the emergence of GenAI.

4.2.2 Machine Learning and Special Education Applications (2022–Early 2023)

With the maturation of machine learning technology, research focus shifted toward Machine Learning and Reinforcement Learning. Notably, keywords such as Autism appeared

during this period, indicating that the early practical direction of AI affective technologies in the educational domain was primarily centered on assisting students with special educational needs in emotion recognition and social skills training. This stage comprised 48 articles (12.0%), indicating that the field began extending from purely technical research toward educational application domains.

4.2.3 Technological Maturation and Diffusion into Educational Settings (2023–2024)

At this stage, research focus gradually shifted toward Emotion Recognition and Education Computing, accompanied by a prominent increase in the Human keyword, indicating that technology had formally entered general educational settings on a large scale beyond purely algorithmic development. Research emphasis shifted from how to train models toward how to apply recognition systems to educational contexts, with growing attention to the role of humans in computing environments. The 2024 publication count reached 105 articles (26.2%), four times the 2022 figure, reflecting the rapid expansion of educational application research.

4.2.4 Interaction and Dialogue with Generative AI (2024–2025)

In the most recent hotspot areas, the emergence of keywords such as Critical Thinking and Psychology Computing marks the full manifestation of GenAI's influence. Research no longer merely identifies emotions but further explores the mutual relationship between generative AI and humans, as well as whether students possess critical thinking skills rather than blindly accepting AI responses. The appearance of Contrastive Learning and Federated Learning indicates that researchers have begun applying different architectures to train generative AI models. The 2025 annual publication count reached 197 articles (49.1%), comprising almost half of the total sample, fully demonstrating the explosive nature of this stage.

Surveying the four evolutionary stages, this study confirms 2023 as the critical inflection point of the field from longitudinal analysis. Unlike early research that described AI primarily as a passive agent relying on sensor-based signal detection (Wei & Li, 2024), the longitudinal data in this study confirm that with the introduction of large language models, AI's role has undergone a fundamental transformation, acquiring the initiative for empathic dialogue (Zhang et al., 2024). This trend is also evident in the exponential growth of annual publication counts: the research focus of affective education has shifted from mere signal processing toward leveraging conversational AI to construct dynamic affective scaffolding that provides more warmly supportive learning.

Figure 4-2.

Longitudinal Overlay Map from Scopus Database

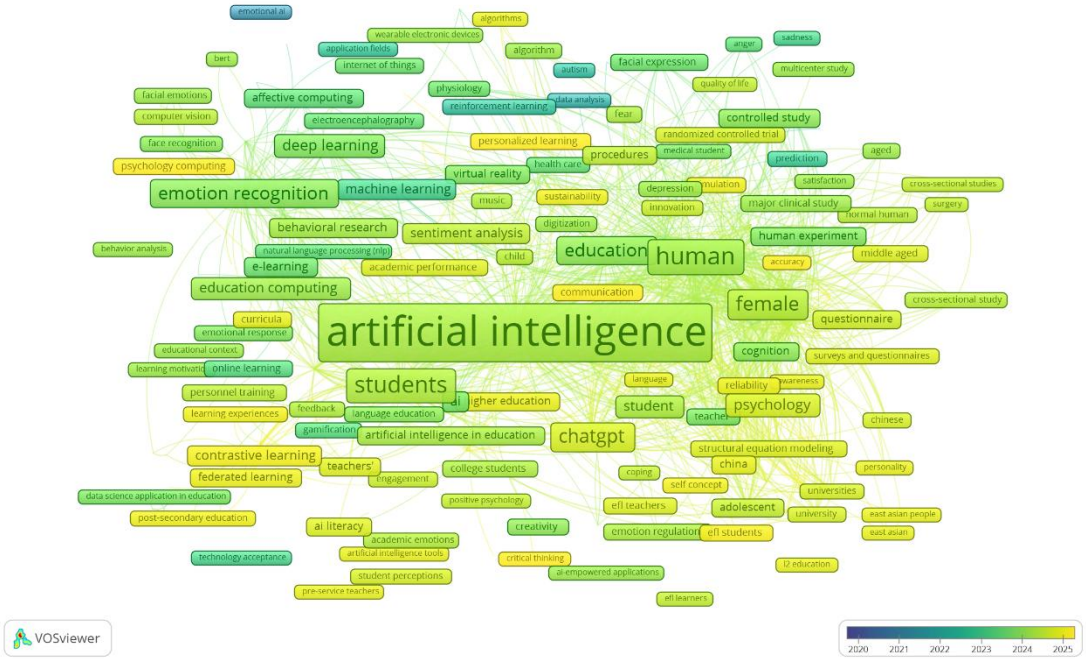


Table 4-2.

Annual Document Distribution

Year	Affective	Emotion	Total	Percentage	Cumulative Percentage
2020	4	9	13	3.2%	3.2%
2021	2	10	12	3.0%	6.2%
2022	5	21	26	6.5%	12.7%
2023	15	33	48	12.0%	24.7%
2024	26	79	105	26.2%	50.9%
2025	86	111	197	49.1%	100.0%
Total	138	263	401	100%	

4.3 Comparison of Core Keyword Focus for Dual Groups

This study further divided the 401 documents into the "Affective" and "Emotion" keyword groups and conducted a comparison through top-ten high-frequency keyword rankings, as

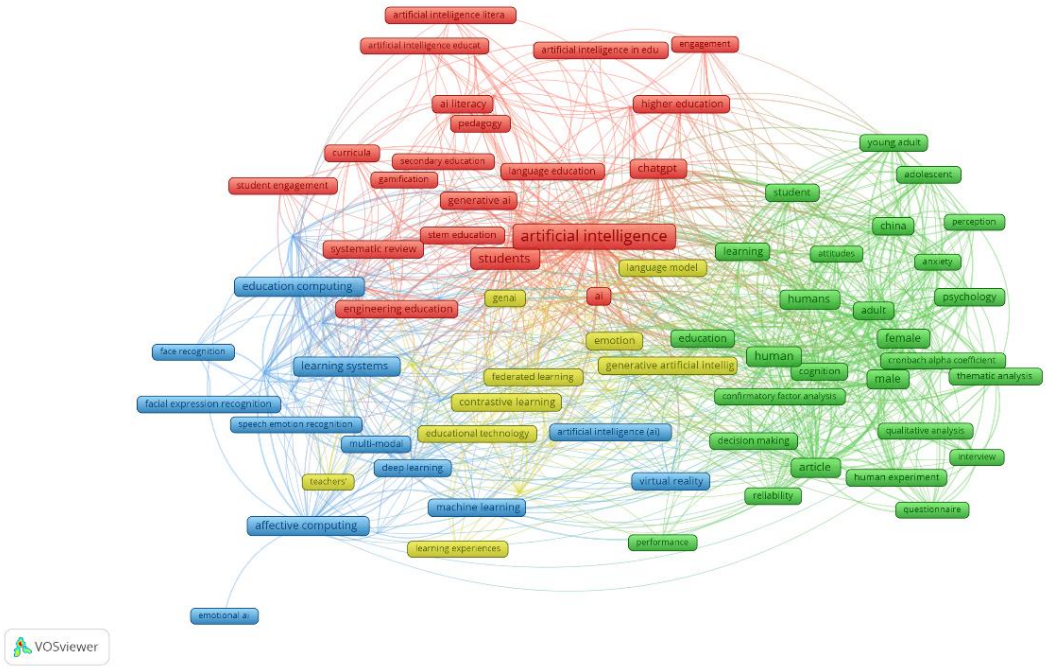
shown in Table 4-3, to explore differences in academic definitions between the two terms in the AI education domain.

Affective Group: System Construction Orientation

The Affective group exhibits pronounced engineering and technical characteristics, as shown in Figure 4-3-1. Beyond the top two keywords shared with the Emotion group — Artificial Intelligence (67 occurrences) and Generative AI (35 occurrences) — Machine Learning (15 occurrences) and Learning Systems (14 occurrences) appear at proportionally higher rates in this group. This indicates that when the academic community focuses on the "Affective" perspective, researchers pay greater attention to algorithmic training approaches, system architectural design, and how to construct a complete affective computing system. In this context, affect is treated as a computable module within a learning system, with research questions oriented toward how to construct.

Figure 4-3-1.

Affective Keyword Network Map from Scopus Database [Figure placeholder]



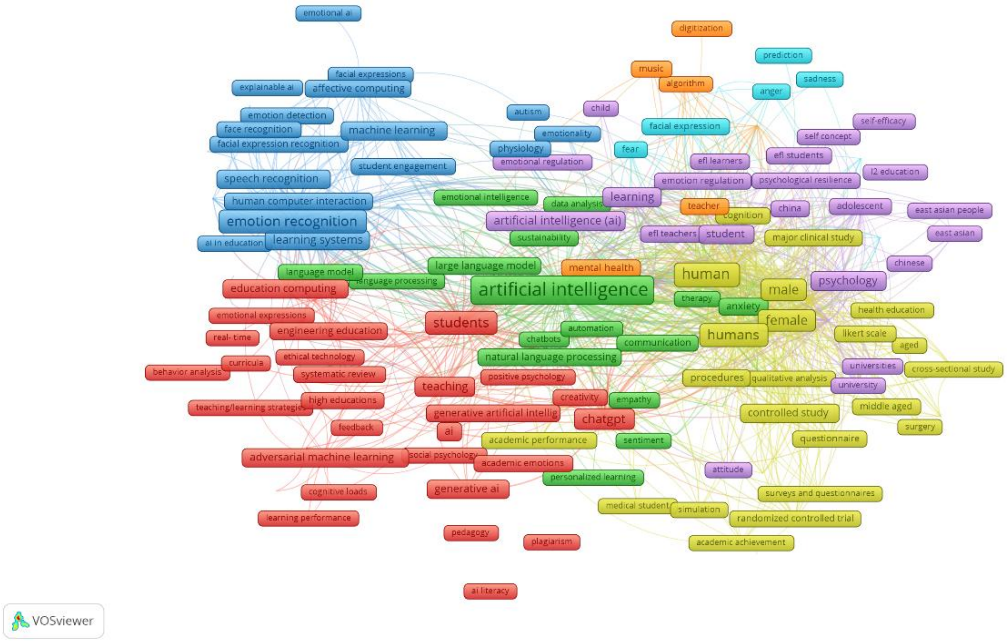
Emotion Group: Recognition Task Orientation

The distinctive high-frequency keyword in the Emotion group is Emotion Recognition (41 occurrences), indicating a greater focus on functional performance. Researchers tend to use "Emotion" to describe specific tasks executed by AI, such as identifying students' emotional states. This reflects that the research emphasis in this group lies in the external application and

recognition effectiveness of technology, rather than underlying algorithmic architecture, with research questions oriented toward how to recognize.

Figure 4-3-2.

Emotion Keyword Network Map from Scopus Database [Figure placeholder]



Cross-Group Appearance of Gender Keywords

Notably, Female and Male appeared among the top ten keywords in both groups (Affective: 13 occurrences each; Emotion: 38 occurrences each), with the Emotion group's frequency nearly three times that of the Affective group. This indicates that AI affective education is shifting from universal models toward greater individualization, with the academic community beginning to pay close attention to algorithmic bias stemming from gender or cultural factors in models. This trend has important implications for the design of AI educational systems: in pursuing recognition accuracy, it is necessary to simultaneously consider the gender and cultural heterogeneity of participants, ensuring that systems provide unbiased and adaptive emotional support to students from diverse backgrounds.

Table 4-3.

Comparison of Top Ten Keywords for Affective and Emotion Groups from Scopus Database

Rank	Affective **Group**	**Freq.**	Emotion **Group**	**Freq.**
1	Artificial Intelligence	67	Artificial Intelligence	152
2	Generative AI	35	Generative AI	71

3	Machine Learning	15	Emotion Recognition	41
4	Learning Systems	14	Female	38
5	Female	13	Male	38
6	Male	13	Machine Learning	37
7	Affective Computing	12	Education	31
8	Education Computing	11	Adult	30
9	Adult	10	Emotions	28
10	Contrastive Learning	10	Teaching	22

In summary, "Affective" and "Emotion" represent two distinct problem orientations in technical implementation: the former focuses on the construction logic of underlying systems, while the latter focuses on the application effectiveness of specific recognition tasks. However, the shared high-frequency keywords in both groups also indicate that, despite their different problem orientations, the two academic communities are jointly responding to educational issues and algorithmic fairness challenges in the GenAI era.

4.4 Five-Dimensional Quantitative Comparison and Analysis

To further validate through quantitative means the differences presented by the keyword comparison above, this study conducted five-dimensional content analysis on the two groups and evaluated the statistical significance of differences using chi-square tests, with results shown in Table 4-4-1. All five dimensions reached statistically significant differences, with effect sizes ranging from Cramer's V = .178 to .300, confirming systematic divergence in research orientations between the two keyword groups rather than random conflation. The following presents key findings for each dimension in turn.

Research Purpose (χ^2=13.26, p=.010, V=.182): Both groups showed P4 "evaluation of existing tool impact" as the largest category — 50.0% in the Affective group and 47.1% in the Emotion group — indicating that current mainstream research focuses on assessing the impact of AI tools on affective perceptions. The most pronounced difference, however, lies in the proportion of P1 "recognition/diagnosis": the Emotion group's 27.4% is nearly double the Affective group's 13.8%, indicating that research using "Emotion" as a keyword tends more toward a technical focus on emotion recognition.

Measurement Channel (χ^2=17.03, p=.004, V=.206): The Affective group heavily relies on M1 traditional questionnaire self-report at 59.4%, indicating a greater emphasis on measuring

learners' subjective perceptions; the Emotion group's M3 facial expression/speech measurement at 21.7% is more than double the Affective group's 9.4%, echoing its technical emphasis on capturing external behaviors. The two groups exhibit a clear philosophical divergence in measurement approach: the former tends toward subjective self-report, and the latter toward objective behavioral capture.

Research Participants (χ^2=35.98, p<.001, V=.300): This dimension shows the most pronounced difference, with an effect size reaching moderate levels. The Affective group has a majority of studies with clearly defined educational targets — undergraduates at 40.6% and K-12 students at 21.7%; the Emotion group shows 55.1% with no specified educational targets, indicating that a large proportion of research using "Emotion" as a keyword essentially constitutes general technical testing, with the educational domain serving merely as a validation platform.

Learning Environment (χ^2=19.66, p=.0002, V=.221): The two groups' distributions are nearly completely reversed. The Affective group's largest category is E1 authentic classroom at 42.0%, while the Emotion group's largest category is E4 unspecified environment at 47.9%, with authentic classroom accounting for only 22.1%. This result closely mirrors the findings on research participants, jointly delineating the starkly different research ecologies of the two keyword groups.

Subject Domain (χ^2=12.77, p=.005, V=.178): Both groups show L4 general/non-specific subject as the largest category, but the Emotion group's proportion at 65.8% exceeds the Affective group's 54.3%. The Affective group's proportion in language learning (L2) at 21.7% is notably higher than the Emotion group's 14.1%, indicating that affective research has deeper application roots in the language education domain.

Table 4-4-1.

Summary Table of Chi-Square Test Results for Dual Groups Across Five Dimensions

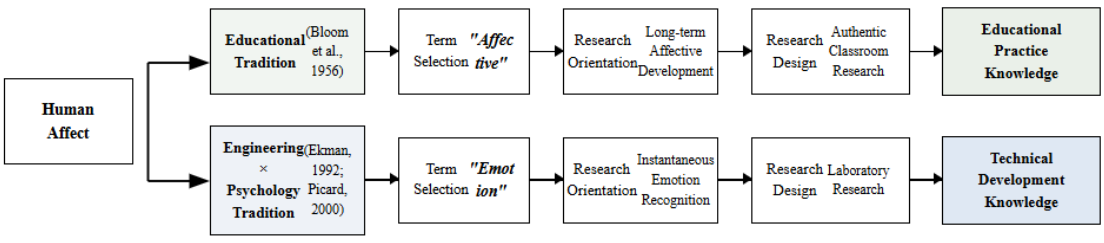
Dimension	χ^2	df	p	Cramer's V
Research Purpose	13.26	4	.010 *	.18
Measurement Channel	17.03	5	.004 * *	.21
Research Participants	35.98	4	<.001 * * *	.30
Learning Environment	19.66	3	<.001 * * *	.22

Subject Domain	12.77	3	.005 * *	.18

Synthesizing the statistical results across five dimensions, the two keyword groups present starkly different research profiles. The typical profile of Affective research is: measuring affective perceptions through questionnaires in authentic classrooms with clearly defined learners — predominantly undergraduates — with research purposes primarily focused on evaluating the impact of existing AI tools, consistent with an educational practice-oriented research logic. The typical profile of Emotion research is: capturing emotional signals through facial expressions or speech in unspecified environments with undesignated participants, with research purposes primarily focused on developing and validating recognition technology, closer to a technology development-oriented engineering research logic. Based on these empirical results, this study inductively proposes the "Dual Knowledge Production Model of Affective and Emotion AI Research": "Affective" is rooted in the educational tradition represented by Bloom et al. (1956), producing knowledge along the pathway of "educational tradition → term selection → long-term affective development → classroom research → educational practice knowledge"; "Emotion" is rooted in the engineering and psychological tradition represented by Ekman (1992) and Picard (2000), producing knowledge along the pathway of "engineering and psychological tradition → term selection → instantaneous emotion recognition → experimental research → technical development knowledge," as depicted in Figure 4-4. This model reinterprets the "terminological conflation" described in existing literature (Wei & Li, 2024; Jiang & Huang, 2023) as a systematic divergence between two academic communities operating under their respective knowledge-production logics, and provides a clear theoretical orientation for subsequent cross-disciplinary collaboration.

Figure 4-4.

Dual Knowledge Production Model of Affective and Emotion AI Research



5. Conclusion and Recommendations

This study employed bibliometric analysis to systematically review the development

trajectory of AI in the affective and emotional education domain in the Scopus database from 2020 to 2025. Synthesizing the analytical results for RQ1 through RQ3, three core conclusions are drawn.

First, cluster analysis shows that the field has developed four mature and well-differentiated research facets: the green cluster establishes the technical foundation in deep learning; the dark blue cluster represents the rigorous measurement of learners' psychological responses through questionnaires and scales in experimental settings; the yellow cluster demonstrates AI's predictive capability in special education and emotional warning applications; and the red cluster grounds these technologies in authentic instructional settings, exploring curriculum design and AI literacy. The formation of this ecosystem indicates that AI affective research has moved beyond the laboratory and is capable of addressing complex problems in actual educational settings.

Second, generative AI has prompted a dramatic shift in research directions. From the annual document distribution, only 51 articles were published from 2020 to 2022 in total, while 2025 alone yielded 197 articles — approximately 15 times the 2020 figure. Prior to 2023, research largely focused on using sensors for passive emotion recognition (Wang et al., 2022; Zhang et al., 2020); however, with the rise of GenAI and conversational AI, the research focus rapidly shifted toward active affective interaction and dialogue generation, representing AI's transformation from a monitoring role to that of a social partner capable of providing emotional support (Pei et al., 2024).

Third, the terminological divergence has a statistical basis: chi-square tests show statistically significant differences across the five research dimensions for the two groups. Research using "Affective" as a keyword exhibits an educational practice orientation, focusing on authentic classrooms with clearly defined educational targets and subjective perception measurement. Research using "Emotion" exhibits a technology development orientation, with more than half failing to specify educational targets and greater reliance on capturing external behaviors such as speech and facial expressions. Based on these empirical results, this study inductively proposes the "Dual Knowledge Production Model of Affective and Emotion AI Research," reinterpreting the terminological conflation described in existing literature (Wei & Li, 2024; Jiang & Huang, 2023) as a systematic divergence between two academic communities operating under their respective knowledge-production logics, and providing a clear theoretical orientation for cross-disciplinary collaboration.

Limitations

This study has the following limitations. First, linguistic and cultural limitations: this study is restricted to English-language literature and may therefore miss localized research published in other languages. Second, terminology standardization limitations: although this study reduces error through synonym merging, some emerging terms in this newest research domain may not have formed discernible trends and thus cannot appear in the knowledge map. Third, analytical scope limitations: bibliometric analysis can examine macro-level structures and trends but cannot conduct meta-analysis of the specific outcomes of each individual study as a systematic literature review can. Fourth, coding method limitations: while the triangulation procedure produced overall acceptable reliability, the low Kappa values for some dimensions in the Emotion group indicate that the conceptual boundaries of this term remain somewhat ambiguous; future research should involve more domain experts for manual coding.

Recommendations

Based on the above findings, this study offers four recommendations for future AI education research. First, as the longitudinal analysis shows the field's research focus has shifted from signal processing toward conversational affective scaffolding, future research should further validate the specific intervention effectiveness of affective scaffolding on learners' emotional regulation, particularly targeting the undergraduate population that constitutes the largest proportion of the Affective group. Second, the prominent appearance of gender-related keywords in both groups indicates that AI affective education is moving from universal models toward individualization; rigorous algorithmic bias verification mechanisms should be established for demographic characteristics such as gender (Sainz-de-Baranda Andujar et al., 2022) to ensure educational equity. Third, given that the five-dimensional statistical results demonstrate the systematic division of labor between the two terms, the academic community should actively build cross-disciplinary collaborative frameworks integrating "Affective" and "Emotion" research, combining computational performance with educational-psychological depth to fill the structural gaps in current research (Yuvaraj et al., 2025). Fourth, the risk of human–AI affective dependency must be taken seriously: against the backdrop of rapidly increasing P3 empathic partner research in the Emotion group, ethical boundaries must be defined and crisis referral mechanisms implemented (Fang et al., 2025) to ensure timely human professional support when necessary.

Declarations Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest: The authors declare no conflict of interest.

Consent to Participate declaration: Not applicable.

Consent to Publish declaration: Not applicable.

Ethics declaration: Not applicable.

Data Availability Statement: Not applicable.

References

Arruda, H., Silva, E. R., Lessa, M., Proença, D., Jr., & Bartholo, R. (2022). VOSviewer and bibliometrix. Journal of the Medical Library Association, 110(3), 392. https://doi.org/10.5195/jmla.2022.1434

Bansal, G., Chamola, V., Hussain, A., Guizani, M., & Niyato, D. (2024). Transforming conversations with AI: A comprehensive study of ChatGPT. Cognitive Computation, 16(5), 2487–2510. https://doi.org/10.1007/s12559-023-10236-2

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain. Longman.

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. Sociological Methods & Research, 42(3), 294–320.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Dilekçi, Ü., & Manap, A. (2022). A bibliometric analysis of research on teacher emotions. International Journal of Psychology and Educational Studies, 9(4), 1222–1235. https://doi.org/10.52380/ijpes.2022.9.4.860

Ekman, P. (1992). Are there basic emotions? Psychological Review, 99(3), 550–553. https://doi.org/10.1037/0033-295X.99.3.550

Elyoseph, Z., Hadar-Shoval, D., Asraf, K., & Lev-Ari, L. (2023). ChatGPT outperforms

humans in emotional awareness evaluations. Frontiers in Psychology, 14, Article 1199058. https://doi.org/10.3389/fpsyg.2023.1199058

Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W., Pataranutaporn, P., ... & Agarwal, S. (2025). How AI and human behaviors shape psychosocial effects of extended chatbot use: A longitudinal randomized controlled study [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2503.17473

Gligorea, I., Cioca, M., Oancea, R., Gorski, A. T., Gorski, H., & Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. Education Sciences, 13(12), 1216. https://doi.org/10.3390/educsci13121216

Jiang, Q., & Huang, Y. (2023, September). Affective computing research progress and its implications for education technology: A bibliometric analysis based on Web of Science via VOSviewer. In 2023 4th International Conference on Artificial Intelligence and Education (ICAIE 2023) (pp. 425–437). Atlantis Press. https://doi.org/10.2991/978-94-6463-242-2_53

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174. https://doi.org/10.2307/2529310

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. Computers and Education: Artificial Intelligence, 2, 100020. https://doi.org/10.1016/j.caeai.2021.100020

Pei, G., Li, H., Lu, Y., Wang, Y., Hua, S., & Li, T. (2024). Affective computing: Recent advances, challenges, and future trends. Intelligent Computing, 3, 0076. https://doi.org/10.34133/icomputing.0076

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. Educational Psychology Review, 18(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Picard, R. W. (2000). Affective computing. MIT Press.

Rao, M. E., & Rao, D. M. (2021). The mental health of high school students during the COVID-19 pandemic. Frontiers in Education, 6, Article 719539. https://doi.org/10.3389/feduc.2021.719539

Sainz-de-Baranda Andujar, C., Gutiérrez-Martín, L., Miranda-Calero, J. Á., Blanco-Ruiz, M., & López-Ongil, C. (2022). Gender biases in the training methods of affective computing: Redesign and validation of the Self-Assessment Manikin in measuring emotions via audiovisual clips. Frontiers in Psychology, 13, Article 955530. https://doi.org/10.3389/fpsyg.2022.955530

Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2023). Human–AI collaboration enables more empathic conversations in text-based peer support. Nature Machine Intelligence, 5(1), 46–57. https://doi.org/10.1038/s42256-022-00593-2

Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., ... & Wen, Q. (2026). Large language models for education: A survey and outlook. IEEE Signal Processing Magazine, 42(6), 51–63.

Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., ... & Zhang, W. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. Information Fusion, 83, 19–52. https://doi.org/10.1016/j.inffus.2022.03.009

Wei, T., & Li, M. (2024). A bibliometric analysis of emotion artificial intelligence in education using VOSviewer and CitNetExplorer. Iris Journal of Education & Research, 3(3). https://doi.org/10.33552/IJER.2024.03.000565

Yuvaraj, R., Mittal, R., Prince, A. A., & Huang, J. S. (2025). Affective computing for learning in education: A systematic review and bibliometric analysis. Education Sciences, 15(1), 65. https://doi.org/10.3390/educsci15010065

Zhang, J., Yin, Z., Chen, P., & Nichele, S. (2020). Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. Information Fusion, 59, 103–126. https://doi.org/10.1016/j.inffus.2020.01.011

Zhang, Y., Yang, X., Xu, X., Gao, Z., Huang, Y., Mu, S., ... & Yu, G. (2024). Affective computing in the era of large language models: A survey from the NLP perspective. [Preprint] arXiv. https://doi.org/10.48550/arXiv.2408.04638