Research

Psikiyatride Güncel Yaklaşımlar-Current Approaches in Psychiatry 2026; 18(Suppl 1):425-441

doi: 10.18863/pgy.2019434

Digital Risks and AI-Enabled Responses in Adolescent Mental Health: A Bibliometric and Science-Mapping Analysis

Adölesan Ruh Sağlığında Dijital Riskler ve Yapay Zekâ Destekli Yanıtlar: Bibliyometrik ve Bilim Haritalama Analizi

Ecem Çiçek[1]

[1]Bartın University, Bartın

ABSTRACT

Objective: This study mapped digital risks and AI-enabled responses in adolescent mental health, quantified publication balance, and tested explicit domain integration within single records.

Method: Scopus metadata (frozen August 9, 2026; exported August 15, 2026) covering articles, reviews, and conference papers were analyzed via bibliometric mapping. Temporal trends evaluated 2015–2025 (complete years), reporting 2026 provisionally. Records were classified as Risk-only, Response-only, or Both. Robustness checks included audits, manual validation of Both records, sensitivity analyses, VOSviewer mapping, thematic evolution, burst detection, and network validation.

Results: Of 4,804 retrieved records, 1,623 formed the primary corpus (1,335 in 2015–2025: 833 Risk-only, 495 Response-only, 7 Both). The response-to-risk ratio rose from 0.04 (2015) to 1.19 (2025), when response output first surpassed risk. Manual review confirmed only 5 direct bridges in 2015–2025 (0.37%) and 9 overall (0.55%); integration remained <1% across sensitivity analyses. Keyword mapping revealed distinct cyberbullying/victimization, AI/digital-mental-health, and machine-learning/suicide-prediction clusters, with recent growth in AI, digital phenotyping, and LLMs.

Conclusion: Although AI response research has expanded rapidly, explicit co-framing of digital risk and AI response within single publications remains rare and independent of 2026 data.

Keywords: School teachers, psychotropic drugs, attitude, mental disorders

ÖZ

Amaç: Bu çalışma, ergen ruh sağlığındaki dijital riskleri ve yapay zekâ destekli yanıtları haritalandırmış, yayın dengesini nicelleştirmiş ve her iki alanın aynı yayınlar içinde ne sıklıkta açıkça entegre edildiğini test etmiştir.

Yöntem: Scopus veri tabanına ait kayıtlar (9 Ağustos 2026'da dondurulan; 15 Ağustos 2026'da dışa aktarılan metadalar), makale, derleme ve bildiri türlerini kapsayacak şekilde bibliyometrik haritalama ile incelenmiştir. Zamansal eğilimler 2015–2025 (tam yıllar) dönemi için değerlendirilmiş, 2026 yılı ise geçici olarak raporlanmıştır. Kayıtlar Sadece Risk, Sadece Yanıt veya Her İkisi şeklinde sınıflandırılmıştır. Dayanıklılık kontrolleri; denetimleri, Her İkisi kategorisindeki tüm kayıtların manuel doğrulanmasını, duyarlılık analizlerini, VOSviewer haritalamasını, tematik gelişimi, patlama tespitini ve ağ doğrulamasını içermiştir.

Bulgular: Elde edilen 4.804 kayıttan 1.623'ü birincil korposu oluşturmuştur (2015–2025 döneminde 1.335 kayıt: 833 Sadece Risk, 495 Sadece Yanıt, 7 Her İkisi). Yanıt/risk oranı 2015'teki 0,04 değerinden, yanıt üretiminin ilk kez riski geçtiği 2025 yılında 1,19'a yükselmiştir. Manuel inceleme, 2015–2025 korposunda yalnızca 5 (%0,37) ve tüm kesitte toplam 9 (%0,55) doğrudan köprü bulunduğunu doğrulamış; entegrasyon oranı duyarlılık analizlerinde %1'in altında kalmıştır. Anahtar kelime haritalaması; siber zorbalık/mağduriyet, YZ/dijital-ruh-sağlığı ve makine-öğrenimi/intihar-tahmini alanlarının belirgin ayrımını, son dönemde ise YZ, dijital fenotipleme ve büyük dil modellerindeki büyümeyi göstermiştir.

Sonuç: Yapay zekâ destekli yanıt araştırmaları hızla büyümüş olsa da, dijital risk ile yapay zekâ yanıtının aynı yayın içinde açıkça bir arada ele alınması nadir görülmekte ve 2026 geçici verilerine dayanmamaktadır.

Anahtar sözcükler: Öğretmenler, psikotrop ilaçlar, tutum, ruhsal bozukluklar

Address For Correspondence: Ecem Çiçek, Bartın University, Faculty of Health Sciences, Department of Public Health Nursing, Bartın, Türkiye
e-mail: egumus@bartin.edu.tr
Received: 15.08.2026 | **Accepted:** 01.10.2026

Introduction

Adolescence is a developmental period in which many mental disorders first emerge and in which social environments can have long-lasting effects on health trajectories. Globally, approximately one in seven people aged 10-19 years experiences a mental disorder, while depression, anxiety, and behavioral disorders remain major causes of illness and disability in this age group (WHO 2025). Digital environments are now part of that developmental context rather than a separate exposure. Recent public-health guidance therefore treats social media, artificial intelligence, and other digital technologies as digital determinants that may simultaneously create opportunities for connection and care and expose vulnerable young people to harm (World Health Organization Regional Office for Europe 2025).

The relationship between digital technology and adolescent mental health is neither uniform nor unidirectional. Online spaces can support peer connection, identity exploration, and access to information, but they also enable cyberbullying, cybervictimization, harassment, social comparison, harmful-content exposure, image-based abuse, and newer forms of synthetic-media manipulation. Cyberbullying has been associated with depressive symptoms, anxiety, self-harm, and suicidal outcomes (Aboujaoude et al. 2015, Hamm et al. 2015, John et al. 2018), while cohort and qualitative studies show that the mental-health meaning of digital engagement depends on intensity, context, social processes, and adolescents' experiences online (Kelly et al. 2018, O'Reilly et al. 2018, Viner et al. 2019). Emerging work on deepfakes further illustrates how technological change can alter the form of interpersonal digital violence (Dalouh Ounia et al. 2025).

Artificial intelligence adds a second and increasingly consequential layer. Machine-learning systems are being developed for screening, classification, suicide-risk prediction, digital phenotyping, and decision support, while conversational agents can provide psychoeducation or low-threshold emotional support (Torous et al. 2021, Dosovitsky and Bunge 2023, Kadirvelu et al. 2026). Evidence is beginning to show potential benefit: a 2025 systematic review and meta-analysis found that artificial-intelligence-driven conversational agents can improve mental-health outcomes among adolescents and young adults, although effects vary by population and intervention characteristics (Feng et al. 2025). At the same time, generative artificial intelligence introduces risks related to privacy, bias, information integrity, crisis management, persuasive interaction, and overreliance (Wang et al. 2025). Evaluations of youth-facing psychotherapy chatbots have raised particular concerns about risk monitoring and crisis handling (Clark 2025, Sobowale et al. 2025). Artificial intelligence therefore occupies a dual position: it can be part of the response to adolescent mental-health problems while also becoming part of the risk environment itself.

Recent syntheses demonstrate rapid growth but also fragmentation. A 2026 bibliometric analysis mapped 1,202 adolescent cyberbullying articles published between 2001 and 2025 and found strong concentration on mental-health outcomes and psychosocial risk, with artificial-intelligence-based detection and digital prevention emerging more recently (Babar et al. 2026). A separate bibliometric study of media manipulation and adolescent mental health identified digital literacy and algorithmic transparency as important policy themes (Pellegrino and Stasi 2024). In parallel, a systematic scoping review of 88 studies on artificial intelligence in adolescent mental-health care found a marked emphasis on diagnostic applications and comparatively limited attention to other stages of care and meaningful end-user involvement (Sharma et al. 2025).

These literatures describe important parts of the problem, but they do not establish whether the risk and response sides of the field are explicitly connected within the same publications. A scientific domain can expand rapidly while its published records remain thematically separated: one stream may document cyberbullying, victimization, and emerging digital harms, whereas another develops prediction algorithms, digital phenotyping, or conversational tools. We therefore define an integration gap as the scarcity of publications that explicitly co-frame digital or artificial-intelligence-mediated risk and artificial-intelligence-enabled response within the same bibliographic record. This is an operational bibliometric construct: it captures limited within-publication co-framing and should not be interpreted as evidence that functional, clinical, interdisciplinary, translational, or cross-publication integration is absent across the broader research ecosystem.

This study therefore aimed to map the global scientific landscape of adolescent digital mental health indexed from 2015 through the provisional 2026 snapshot; quantify the balance and temporal relationship between risk-focused and artificial-intelligence-enabled response research using complete calendar years through 2025 for trend inference; identify influential countries, sources, authors, and publications; characterize conceptual, intellectual, geographic, and collaboration structures; examine thematic evolution and emergent keyword bursts; and test whether the apparent integration gap remains under record-level audit, manual bridge validation, and alternative corpus definitions.

Method

Study Design and Analytic Framework

A bibliometric and science-mapping design was used to quantify publication performance, conceptual structure, intellectual structure, collaboration patterns, and temporal thematic change. The workflow followed established recommendations for transparent corpus construction and science mapping (Donthu et al. 2021) and VOSviewer methodology (van Eck and Waltman 2010). Reporting was checked against the BIBLIO minimum requirements for bibliometric reviews of biomedical literature (Montazeri et al. 2023).

The bibliographic record was the unit of analysis. Descriptive science-mapping analyses used the full indexed snapshot, whereas inferential statements about temporal change were anchored to complete calendar years (2015-2025) and treated 2026 separately as provisional.

Data Source, Time Window, and Search Strategy

Scopus was selected as the sole bibliographic source to maintain a single internally consistent framework for citation counts, cited references, affiliations, author identifiers, and conference-proceedings metadata across psychiatry, psychology, public health, education, and computer science. Scopus provides curated multidisciplinary bibliometric metadata and substantial proceedings coverage, which is relevant to rapidly evolving artificial-intelligence research (Baas et al. 2020). This choice was made for internal consistency rather than as a claim of exhaustive coverage; comparative studies show that bibliographic databases differ in journal and disciplinary coverage (Singh et al. 2021).

The search syntax, concept blocks, publication-year filter, and eligibility rules were frozen on 9 August 2026. The same protocol was exported from Scopus on 15 August 2026, which is the metadata snapshot used in the analysis. The executed query restricted publication years to 2015-2026 (PUBYEAR > 2014 AND PUBYEAR < 2027) and applied no language restriction. Because 2026 was incomplete at extraction, 2015-2025 served as the primary interval for temporal inference; 2026 records were retained for the cross-sectional bibliometric snapshot and reported separately as provisional. The exact Scopus query is reproduced in Supplementary Table S1.

The search combined three concept blocks: (1) adolescent population terms; (2) explicit mental-health outcomes, including depression, anxiety, suicidality, self-harm, psychological distress, loneliness, and well-being; and (3) either digital-risk terms such as cyberbullying, cybervictimization, online harassment, image-based abuse, deepfakes, artificial-intelligence companions, or algorithmic harms, or artificial-intelligence technology terms combined with response-action terms such as detection, prediction, screening, prevention, intervention, support, monitoring, or moderation. The search was deliberately broader than the final analytic corpus so that relevance refinement could be performed transparently after retrieval.

Document Eligibility, Deduplication, and Relevance Refinement

Articles, reviews, and conference papers were eligible. Conference papers were retained because conference proceedings are an important dissemination route in computing and artificial-intelligence research; excluding them could systematically underrepresent rapidly evolving technical work. Editorials, notes, letters, book chapters, short surveys, and other document types were excluded. Duplicate records

were identified by DOI when available and by normalized title otherwise; when duplicate metadata were encountered, the more highly cited record was retained.

Because a broad title-abstract-keyword query can retrieve records in which a target term is incidental, a deterministic rule-based relevance refinement was applied to define a high-specificity analytic corpus. An adolescent term had to occur in the title or author keywords; a mental-health term had to occur in the title or abstract; and the record had to contain either a predefined digital/artificial-intelligence-mediated risk term in the title or abstract, or an artificial-intelligence term in the title or author keywords together with a response-action term in the title or abstract. Exclusions were applied sequentially and assigned to the first failed rule: absence of a central adolescent term, absence of an explicit mental-health term, or absence of a qualifying risk/AI-response configuration. The dictionaries, Boolean logic, and exclusion cascade are provided in Supplementary Tables S2-S3.

To evaluate possible rule-based misclassification, a post hoc record-level audit used a stratified random sample of 60 excluded records (allocated proportionally across the three exclusion rules) and a simple random sample of 60 retained records, both generated with seed 20260906. Titles, abstracts, and author keywords were manually reviewed; potential false exclusions, borderline exclusions, and peripheral/potential false retentions were recorded, with Wilson 95% confidence intervals for the principal audit rates (Supplementary Table S4).

Risk-Response Classification and Integration Gap

Each eligible record was assigned one mutually exclusive bibliographic class using precedence rules. Records meeting both definitions were labeled Both; among the remaining records, those meeting only the risk definition were labeled Risk-only and those meeting only the response definition were labeled Response-only. Risk required a predefined digital or artificial-intelligence-mediated risk term in the title or abstract. Response required an artificial-intelligence term in the title or author keywords plus at least one response-action term in the title or abstract. These classes represent explicit bibliographic framing, not clinical diagnoses, causal categories, or evidence of intervention effectiveness.

For each publication year, a response-to-risk ratio was calculated as (Response-only + Both)/(Risk-only + Both). Values below 1 indicate greater risk-focused publication output, whereas values above 1 indicate greater artificial-intelligence-enabled response output. The deterministic integration share was the proportion classified as Both. Accordingly, the operationalized integration gap measures explicit within-record bibliographic co-framing only. It does not quantify functional integration across clinical or prevention pathways, interdisciplinary collaboration, translational exchange, knowledge transfer between separate research communities, or integration distributed across different publications.

Because this bibliographic integration gap was the study's central finding, all 15 deterministic Both records in the full snapshot were manually reviewed using a stricter direct-bridge criterion: the artificial-intelligence method or system had to function as detection, prediction, prevention, education, safety, or support directed at the same digital/AI-mediated harm identified in the record. Background co-mentions and cases in which AI was only the exposure did not qualify as direct bridges. The deterministic classification remained the reproducible primary metric; manual validation was reported as an additional robustness check (Supplementary Table S8).

Bibliometric Performance and Conceptual-Structure Analysis

Performance indicators included annual publication output, compound annual growth from 2015 to 2025, total and average Scopus cited-by counts, median citations, corpus h-index, document type, language, leading sources, productive authors, and country output. Country productivity used full counting, so a publication contributed one count to each represented country. Internationally coauthored publications were classified as multiple-country publications. To evaluate whether geographic conclusions depended on counting method, country publication and citation ranks were recalculated using fractional allocation across participating countries and compared with full-counting ranks (Supplementary Table S13).

Science mapping was performed in VOSviewer Online (van Eck and Waltman 2010). The main conceptual map used author-keyword co-occurrence with full counting, a minimum occurrence threshold of 10, and association-strength normalization. Additional networks used author co-authorship with fractional counting and at least three documents per author, country citation with fractional counting and at least five documents and 10 citations per country, and cited-author co-citation with full counting and at least 20 citations. Organization-level co-authorship was not retained in the final analytic set because the resulting network was too sparse for stable interpretation.

Obvious lexical variants in author keywords were normalized for secondary quantitative analyses. To quantify map representativeness and structural coherence, the number of records contributing author keywords, retained nodes, and non-zero links was calculated; an independent Louvain community analysis was then applied to the association-strength keyword graph with seed 42, and modularity Q and the within-community share of edge weight were reported (Blondel et al. 2008).

Thematic evolution was summarized for selected normalized keywords across 2015-2019, 2020-2022, and 2023-2025. Keyword burst intervals were identified in complete years 2015-2025 using Kleinberg's burst-detection framework (s=2, gamma=1) for keywords with at least 10 complete-year occurrences (Kleinberg 2002). These secondary analyses were used to describe temporal concentration and network structure, not to infer causality.

Sensitivity Analysis

Two complementary corpus-definition sensitivity analyses were performed using complete years 2015-2025. The nested high-specificity analysis required all primary-corpus criteria and additionally required the mental-health concept in the title or author keywords, together with either a central risk term or a central artificial-intelligence term plus a response action. The expanded-lexicon recall analysis broadened youth terminology, mental-health concepts, digital-risk expressions, artificial-intelligence methods, and response-action terms to test whether the main conclusion resulted from an overly narrow dictionary.

Convergence was evaluated using three explicit criteria: (1) the Both share remained below 1%; (2) the response-to-risk ratio increased from 2024 to 2025; and (3) the ratio exceeded 1 in 2025. The precise crossover year was not required to be identical across corpus definitions.

Software and Reproducibility

Data processing and computational analyses were performed in Python 3.13.5 using pandas 2.2.3 and NumPy 2.3.5. Matplotlib 3.10.8 was used for non-network visualizations, NetworkX 3.6.1 for Louvain community analysis, and VOSviewer Online for bibliometric network construction and visualization. Detailed search strategies, classification dictionaries, sensitivity and robustness analyses, temporal analyses, network-validation procedures, and VOSviewer settings are provided in the Supplementary Material. Reproducible analysis code and derived non-proprietary outputs were retained with the study files.

All search queries, raw dataset files, and visualization parameters used in this bibliometric analysis and listed as supplementary files in the text have been made publicly available on the Zenodo repository to ensure reproducibility. All supplementary tables, figures, and queries mentioned in the text are available at https://doi.org/10.5281/zenodo.23296739.

Ethical Considerations

The study used publicly indexed bibliographic metadata and did not involve recruitment, intervention, access to private clinical records, or analysis of identifiable participant-level data. Ethics committee approval and informed consent were therefore not applicable.

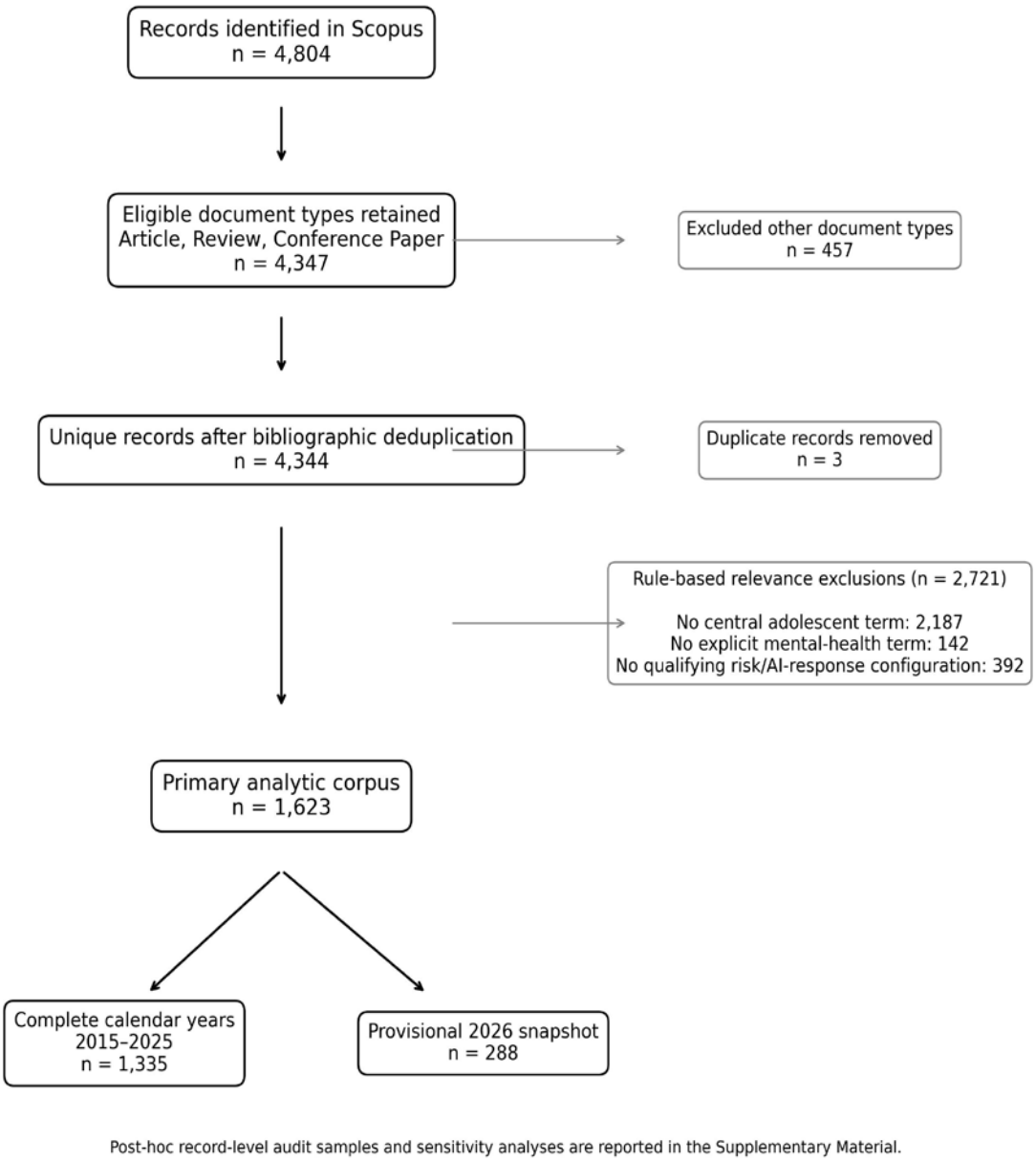


Figure 1. Study selection, sequential exclusion cascade, and analytic time frame.

Rule-based exclusions are shown by first failed criterion. The primary corpus (n=1,623) was separated into complete calendar years 2015-2025 (n=1,335), which were used for temporal inference, and a provisional 2026 snapshot (n=288). Record-level audit samples and sensitivity analyses are reported in the Supplementary Material.

Results

Study Selection and Corpus Characteristics

The Scopus search retrieved 4,804 records. Restriction to articles, reviews, and conference papers retained 4,347 records; deduplication removed three, leaving 4,344 unique records. Sequential relevance refinement excluded 2,187 records without a central adolescent term, 142 without an explicit mental-health term in the title or abstract, and 392 without a qualifying risk or AI-response configuration, yielding a primary analytic corpus of 1,623 (Figure 1; Supplementary Table S3). Of these, 1,335 belonged to complete calendar years 2015-2025 and 288 to the provisional 2026 snapshot. Articles represented 1,374 records (84.7%), conference papers 128 (7.9%), and reviews 121 (7.5%); 1,557 records (95.9%) were in English. In the record-level audit, 55 of 60 sampled exclusions were confirmed, 3 (5.0%; Wilson 95% CI 1.7-13.7%) were

judged potential false exclusions under a broader semantic interpretation, and 2 were borderline because the indexed record did not establish adolescent-specific centrality. Among 60 retained records, 54 were confirmed and 6 (10.0%; 95% CI 4.7-20.1%) were peripheral/potential false retentions under a stricter semantic interpretation (Supplementary Table S4). The full 1,623-record snapshot accumulated 32,053 Scopus citations at extraction, with a mean of 19.75 and a median of 5 citations per record; the corpus h-index was 81. These values are Scopus-specific snapshot indicators and were not normalized for publication age or disciplinary field (Table 1).

Table 1. Characteristics of the primary analytic corpus

Indicator	Value
Records retrieved from Scopus	4,804
Eligible document types before deduplication	4,347
Unique records after deduplication	4,344
Primary analytic corpus, full snapshot	1,623
Complete-year primary subset, 2015-2025	1,335
Provisional 2026 subset	288
Articles	1,374 (84.7%)
Conference papers	128 (7.9%)
Reviews	121 (7.5%)
English-language records	1,557 (95.9%)
Records with at least one author keyword	1,534 (94.5%)
Total Scopus citations, full snapshot	32,053
Mean citations per record	19.75
Median citations per record	5
Corpus h-index	81
Publications in 2015	24
Publications in 2025	377
2015-2025 CAGR	31.7%
Complete-year Risk-only	833 (62.4%)
Complete-year Response-only	495 (37.1%)
Complete-year Both	7 (0.52%)
Manually validated direct bridges, complete years	5 (0.37%)
Deterministic Both, full 2015-2026 snapshot	15 (0.92%)
Manually validated direct bridges, full snapshot	9 (0.55%)

CAGR: compound annual growth rate

Publication Growth and the Risk-Response Transition

Across complete years, annual output increased from 24 publications in 2015 to 377 in 2025, corresponding to a compound annual growth rate of 31.7% (Figure 2). The 2026 snapshot contained 288 records but was not used to infer annual growth or decline because it represented an incomplete observation year. It is shown separately only to describe the current indexed snapshot.

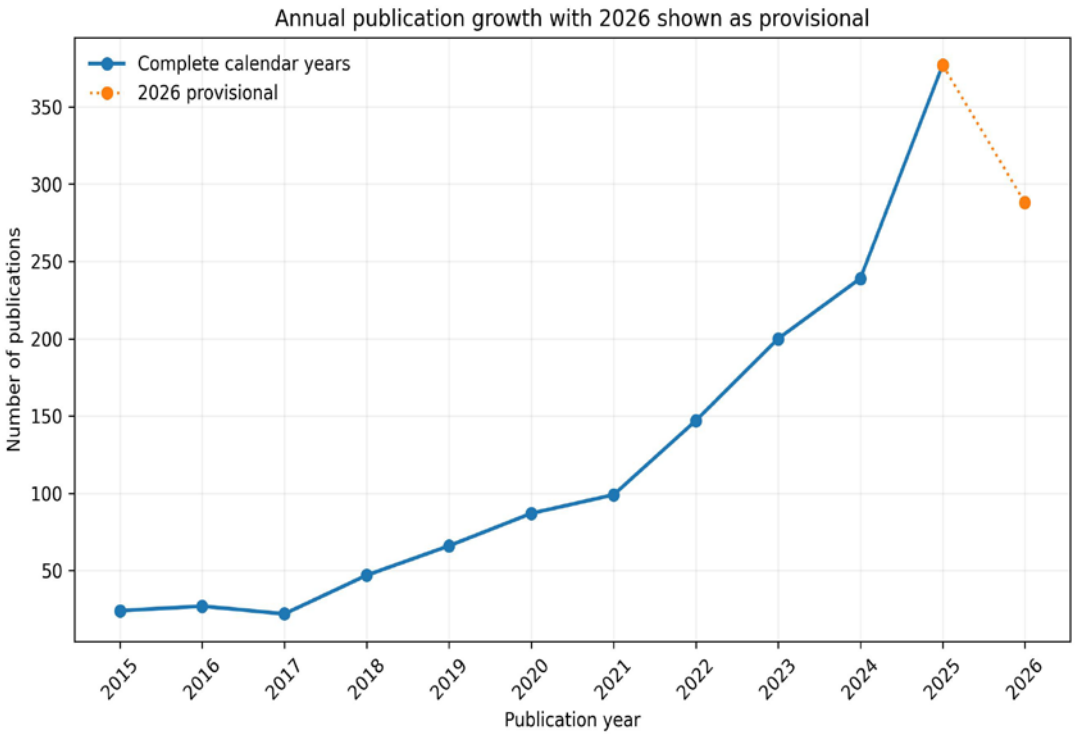


Figure 2. Annual growth of adolescent digital mental-health research.

Solid lines represent complete calendar years 2015-2025. The dotted 2025-2026 extension shows the provisional 2026 snapshot and is not used to infer annual growth or decline.

The complete-year corpus (n=1,335) contained 833 Risk-only records, 495 Response-only records, and 7 Both records. The response-to-risk ratio rose from 0.04 in 2015 to 0.25 in 2019, 0.46 in 2022, 0.88 in 2024, and 1.19 in 2025; thus, response-focused output first exceeded risk-focused output in a complete calendar year in 2025 (Figure 3; Supplementary Table S5). The deterministic Both share in complete years was 0.52% (7/1,335). The provisional 2026 subset contained 117 Risk-only, 163 Response-only, and 8 Both records (ratio 1.37), but these values were not used to establish the crossover or trend (Supplementary Table S6). Across the full snapshot, the descriptive class counts were 950 Risk-only, 658 Response-only, and 15 Both.

Table 2. Leading sources and productive authors

Rank	Source	n	Author	n
1	Journal of Affective Disorders	62	Wright, Michelle F.	13
2	International Journal of Environmental Research and Public Health	55	Rey, Lourdes	11
3	Frontiers in Psychiatry	39	Zhang, Wei	11
4	Frontiers in Psychology	30	Baiden, Philip	10
5	Journal of Medical Internet Research	25	Hermens, Daniel F.	10
6	BMC Public Health	22	Quintana-Orts, Cirenia	9
7	Children and Youth Services Review	20	Hinduja, Sameer	8
8	Computers in Human Behavior	19	Calvete, Esther	8
9	Psychiatry Research	18	Estévez, Estefanía	8
10	Frontiers in Public Health	18	Hong, Jun Sung	8

Geographic, Source, and Author Structure

The most productive source was the Journal of Affective Disorders (n=62), followed by the International Journal of Environmental Research and Public Health (n=55), Frontiers in Psychiatry (n=39), Frontiers in Psychology (n=30), and the Journal of Medical Internet Research (n=25). Michelle F Wright was the most productive disambiguated author in the analytic corpus (n=13), followed by Lourdes Rey and Wei Zhang (n=11 each), and Philip Baiden and Daniel F Hermens (n=10 each) (Table 2).

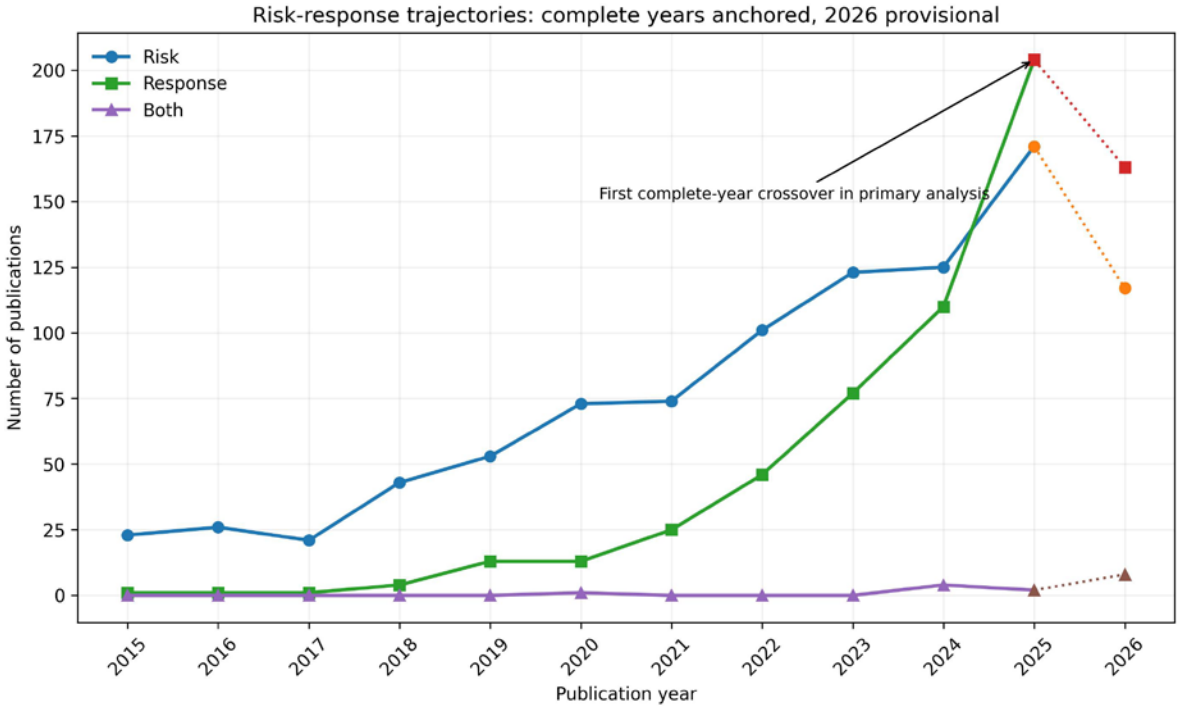


Figure 3. Risk-only, artificial-intelligence-enabled Response-only, and Both publication trajectories.

Solid lines represent complete calendar years 2015-2025; the dotted extension shows provisional 2026. The annotation marks the first complete-year crossover in the primary analysis, when response output exceeded risk output in 2025. Both records meet both deterministic definitions and are classified separately from Risk-only and Response-only records.

Table 3. Leading countries and international collaboration

Rank	Country	Publications	MCP	SCP	MCP ratio
1	United States	444	181	263	0.41
2	China	334	105	229	0.31
3	United Kingdom	144	103	41	0.72
4	Spain	139	50	89	0.36
5	Australia	102	61	41	0.60
6	India	76	9	67	0.12
7	Germany	70	46	24	0.66
8	Canada	67	34	33	0.51
9	South Korea	67	30	37	0.45
10	Italy	58	23	35	0.40
11	Turkey	44	13	31	0.30
12	Hong Kong	43	26	17	0.60
13	Sweden	36	27	9	0.75
14	Portugal	29	17	12	0.59
15	France	28	23	5	0.82

MCP: multiple-country publication; SCP: single-country publication

Using full counting, the United States led publication output with 444 records, followed by China (334), the United Kingdom (144), Spain (139), and Australia (102). International collaboration varied substantially: the multiple-country publication ratio was 0.72 for the United Kingdom, 0.66 for Germany, 0.60 for Australia, and 0.41 for the United States, compared with 0.12 for India; Turkey contributed 44 publications, 13 of which were internationally coauthored (Table 3).

Geographic conclusions were robust to counting method. The United States remained rank 1 for both full and fractional publication allocation and for both full and fractional citation allocation; China, Spain, and the United Kingdom also remained among the leading countries, although some secondary ranks shifted (Supplementary Table S13).

Table 4. Ten most cited publications in the analytic corpus

Rank	First author (year)	Title	Source	Cited by
1	Kelly, Yvonne (2018)	Social Media Use and Adolescent Mental Health: Findings From the UK Millennium Cohort Study	EClinicalMedicine	563
2	John, Ann (2018)	Self-harm, suicidal behaviours, and cyberbullying in children and young people: Systematic review	Journal of Medical Internet Research	439
3	Hamm, Michele P. (2015)	Prevalence and effect of cyberbullying on children and young people: A scoping review of social media studies	JAMA Pediatrics	392
4	Aboujaoude, Elias (2015)	Cyberbullying: Review of an old problem gone viral	Journal of Adolescent Health	381
5	O'Reilly, Michelle (2018)	Is social media bad for mental health and wellbeing? Exploring the perspectives of adolescents	Clinical Child Psychology and Psychiatry	314
6	Armitage, Richard (2021)	Bullying in children: Impact on child health	BMJ Paediatrics Open	281
7	Bottino, Sara Mota Borges (2015)	Cyberbullying and adolescent mental health: Systematic review; [Cyberbullying e saúde mental dos adolescentes: Revisão sistemática]; [Acoso cibernético y la salud mental de los adolescentes: Revisión sistemática]	Cadernos de Saude Publica	274
8	Radovic, Ana (2017)	Depressed adolescents' positive and negative use of social media	Journal of Adolescence	269
9	Boer, Maartje (2021)	Social media use intensity, social media use problems, and mental health among adolescents: Investigating directionality and mediating processes	Computers in Human Behavior	262
10	Viner, Russell M. (2019)	Roles of cyberbullying, sleep, and physical activity in mediating the effects of social media use on mental health and wellbeing among young people in England: a secondary analysis of longitudinal data	The Lancet Child and Adolescent Health	261

Citation counts are Scopus cited-by counts at retrieval and are not field- or age-normalized

Influential Publications

The most highly cited record was Kelly et al. (2018), "Social Media Use and Adolescent Mental Health: Findings From the UK Millennium Cohort Study" (563 citations). It was followed by the systematic review of self-harm, suicidal behaviors, and cyberbullying by John et al. (2018) (439 citations), the scoping review of cyberbullying prevalence and effects by Hamm et al. (2015) (392 citations), and the clinical review of cyberbullying by Aboujaoude et al. (2015) (381 citations). The leading cited literature therefore reflects the field's earlier emphasis on social media, cyberbullying, and adverse mental-health outcomes (Table 4).

Conceptual Structure

Author keywords were available for 1,534 of 1,623 records (94.5%). After normalization, 100 keywords met the minimum occurrence threshold of 10 and contributed 1,540 non-zero co-occurrence links to the association-strength network (Figure 4; Supplementary Table S12). The VOSviewer map showed a dense but differentiated conceptual structure. A risk region was centered on cyberbullying, bullying,

cybervictimization, victimization, aggression, and life satisfaction. A technology-response region connected mental health with artificial intelligence, natural language processing, chatbots, large language models, digital health, and digital phenotyping. Machine learning was strongly associated with suicidal ideation, suicide risk, prediction models, random forest, and non-suicidal self-injury. Social media, school, prevalence, sexting, and well-being formed a contextual region, while adolescent, adolescence, anxiety, and suicide connected multiple areas. Independent Louvain analysis of the same association-strength graph identified four communities with modularity Q=0.292; 61.3% of association-strength edge weight lay within communities. This quantitative validation supports meaningful network structure but was not treated as a one-to-one validation of VOSviewer color clusters.

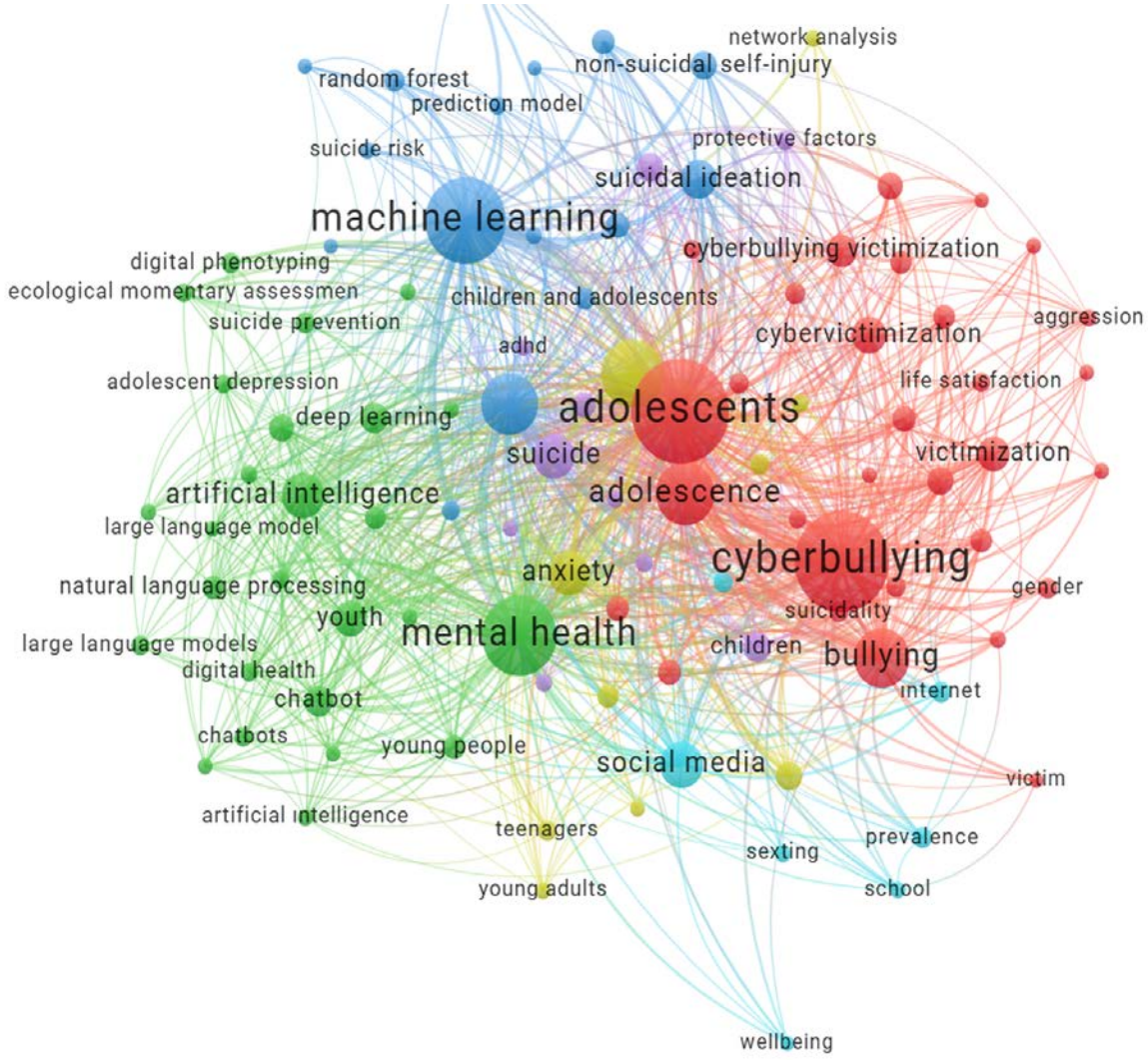


Figure 4. VOSviewer author-keyword co-occurrence network in the primary analytic corpus (n=1,623).

Author keywords were present in 1,534 records (94.5%). Full counting, association-strength normalization, and a minimum occurrence threshold of 10 were used; 100 keywords met the threshold. Node size reflects keyword weight, links and proximity indicate co-occurrence relatedness, and colors represent VOSviewer structural groupings. Quantitative network-validation metrics are reported in Supplementary Table S12.

The supplementary science maps reinforced this structural picture. In the country citation network, the United States occupied the most central position, with China, the United Kingdom, Spain, Canada, Australia, and other countries forming connected secondary hubs (Supplementary Figure S4). The cited-author co-citation network showed several intellectual communities rather than a single homogeneous knowledge base, including traditions associated with bullying/peer victimization and others oriented toward suicidality, psychopathology, and prediction (Supplementary Figure S5). Author co-authorship was more fragmented, with multiple local collaboration clusters and limited intercluster connectivity (Supplementary Figure S6).

Table 5. Interpretive thematic domains visible in the VOSviewer author-keyword co-occurrence map			
Domain	**Interpretive label**	**Representative terms**	**Interpretation**
1	Cyberbullying and victimization	cyberbullying; cybervictimization; bullying; victimization; aggression; life satisfaction; gender	Established peer-to-peer digital harm and psychosocial consequence literature
2	Artificial intelligence-enabled digital mental health	mental health; artificial intelligence; natural language processing; chatbot; large language models; digital health; digital phenotyping	Technology-enabled screening, support, and digital mental-health applications
3	Predictive analytics and suicide risk	machine learning; suicidal ideation; suicide risk; prediction model; random forest; non-suicidal self-injury	Prediction and classification of high-stakes psychiatric outcomes
4	Core adolescent psychopathology bridge	adolescents; adolescence; anxiety; suicide; protective factors; youth	Population and psychiatric outcome concepts connecting multiple regions
5	Social-media and school context	social media; school; prevalence; sexting; well-being; internet	Digital context, school exposure, and population-level well-being

The table summarizes dominant interpretive regions visible in the VOSviewer map and does not constitute a clinical taxonomy. Author keywords were available for 1,534/1,623 records (94.5%); 100 normalized keywords met the >=10-occurrence threshold and formed 1,540 non-zero links. Independent Louvain validation of the association-strength graph identified four communities (Q=0.292), with 61.3% of edge weight within communities; these communities were used as a quantitative structural check rather than as a one-to-one substitute for the VOSviewer color groupings.

Risk Profile and Artificial-Intelligence Emergence

The operationalized risk literature remained overwhelmingly anchored in cyberbullying and cybervictimization: 926 of 965 risk-containing publications (96.0%) contained cyberbullying, cybervictimization, or online-bullying terminology. Explicitly emerging risk terms were uncommon within the predefined lexicon: image-based abuse or deepfakes occurred in six risk publications, artificial-intelligence companions in four, and algorithmic-harm or manipulation terminology in one. These values indicate bibliographic visibility of explicit terms rather than prevalence of real-world exposure.

Artificial-intelligence-related terminology accelerated strongly in complete years after 2022. Machine-learning records increased from 34 in 2022 to 58 in 2023, 79 in 2024, and 129 in 2025; explicit artificial-intelligence terminology increased from 21 in 2024 to 64 in 2025. Large-language-model/generative-AI terminology first appeared in the annual series in 2024 (n=6) and reached 30 in 2025; chatbot/conversational-agent terminology increased from 15 to 31 over the same interval (Supplementary Figure S1; Supplementary Table S9).

Thematic evolution based on normalized author keywords showed that cyberbullying remained common but declined in relative representation from 48.4% of records in 2015-2019 to 21.7% in 2023-2025, whereas machine learning rose from 5.4% to 23.2%, artificial intelligence from 0.5% to 8.1%, and mental health from 12.9% to 20.0% (Supplementary Figure S2; Supplementary Table S10). Exploratory burst detection identified recent bursts for digital phenotyping in 2023-2024 and for artificial intelligence, large language models, and adolescent mental health in 2025 (Supplementary Figure S3; Supplementary Table S11).

Response Functions and Explicitly Integrated Research

Within the 673 Response or Both records, overlapping technology coding identified machine learning in 440 records, artificial intelligence in 158, chatbot or conversational-agent terminology in 104, large language models or generative artificial intelligence in 64, deep learning in 60, natural language processing in 51, and digital phenotyping in 23. Overlapping response-function coding most frequently identified detection or classification (n=485) and prediction or risk assessment (n=411), followed by support or treatment (n=402) and prevention or intervention (n=377). These categories were non-exclusive and indicate thematic visibility rather than intervention prevalence or effectiveness (Supplementary Table S5).

The deterministic algorithm classified 15 records as Both across the full snapshot: one in 2020, four in 2024, two in 2025, and eight in provisional 2026. Manual review using the stricter direct-bridge criterion confirmed 9 of the 15 as true direct bridges. Within complete years, 7 records were algorithmic Both and 5 were manually validated direct bridges, corresponding to 0.37% of the 1,335-record complete-year corpus.

All five validated direct bridges in complete years appeared in 2024-2025, but this very small count was considered insufficient to infer a stable emergence trend. Four additional validated direct bridges occurred in the provisional 2026 subset. The validated records primarily involved machine learning or AI for cyberbullying detection/prediction, online-safety education/support, or precision prevention; records in which cyberbullying was background context or AI was only the exposure were not considered direct bridges (Supplementary Table S8).

Sensitivity Analysis

Both complete-year sensitivity analyses converged on the central conclusion (Supplementary Table S7). The nested high-specificity corpus contained 734 records: 372 Risk-only, 359 Response-only, and 3 Both (0.41%); its response-to-risk ratio was 1.66 in 2024 and 2.39 in 2025. The expanded-lexicon recall corpus contained 1,532 records: 941 Risk-only, 579 Response-only, and 12 Both (0.78%); its ratio increased from 0.89 in 2024 to 1.28 in 2025. Thus, in the primary, high-specificity, and expanded-recall analyses, the Both share remained below 1%, the response-to-risk ratio increased from 2024 to 2025, and the ratio exceeded 1 in 2025. The exact crossover year varied with stringency, but the integration-gap conclusion did not.

Discussion

This study identifies a robust transition in the bibliographic organization of adolescent digital mental-health research. When temporal inference was restricted to complete years, output increased rapidly through 2025 and the balance shifted from clearly risk-dominant research toward greater AI-enabled response output, with the primary response-to-risk ratio first exceeding 1 in 2025. Yet the shift in volume was not matched by explicit within-publication co-framing: only 7 of 1,335 complete-year records met the deterministic Both definition (0.52%), and stricter manual review confirmed only 5 direct bridges (0.37%). The same substantive conclusion held under both a high-specificity subset and an expanded-lexicon recall analysis, each with integration below 1%. The central bibliometric finding is therefore persistent separation in how harm characterization and AI-enabled response are jointly framed within individual publications. This finding concerns the structure of the published record and does not demonstrate that functional, clinical, interdisciplinary, or cross-publication integration is absent across the broader research ecosystem.

The risk architecture remains historically anchored in cyberbullying and cybervictimization. This is consistent with the citation profile, in which foundational reviews and cohort studies on cyberbullying, social media, depression, self-harm, and suicidality remain highly influential (Aboujaoude et al. 2015, Hamm et al. 2015, John et al. 2018, Kelly et al. 2018). It also converges with Babar et al. (2026), who mapped a large cyberbullying literature in which mental-health outcomes and psychosocial risks were central while AI-based detection and digital prevention were more recent themes. The present thematic-evolution analysis adds a temporal dimension: cyberbullying remained visible in absolute terms but declined markedly as a share of period-specific records, while machine learning, artificial intelligence, and mental-health terminology expanded. This pattern should be interpreted as diversification of the research landscape, not disappearance of cyberbullying as a public-health problem.

The response literature has a different center of gravity. In the keyword network, machine learning was positioned near suicidal ideation, suicide risk, prediction models, and non-suicidal self-injury, while artificial intelligence, natural language processing, chatbots, large language models, digital health, and digital phenotyping formed a connected response-oriented region around mental health. The recent bursts for digital phenotyping, artificial intelligence, and large language models further indicate that the technical vocabulary of the field is changing rapidly. This configuration is compatible with Sharma et al. (2025), who found that AI applications in adolescent mental-health care were concentrated heavily in diagnostic

processes, with comparatively less work on other stages of care and end-user involvement. It also contextualizes emerging intervention evidence: conversational-agent meta-analysis suggests potential improvement in some youth mental-health outcomes, but such evidence does not by itself establish safe system-level integration (Feng et al. 2025).

The bibliometric separation between prediction-oriented and harm-oriented literatures has practical implications, but those implications must be distinguished from what the bibliographic data directly demonstrate. The present study shows co-occurrence structure, publication trajectories, and explicit co-framing; it does not test whether an algorithm prevents harm, improves referral, or benefits patients. When interpreted together with external clinical and public-health evidence, however, the pattern highlights an important implementation question: detection is not prevention, and prediction is not protection. A model may identify cyberbullying, depressive language, or suicide risk, yet population benefit depends on downstream actions such as timely human support, school and family response protocols, service capacity, platform-level exposure reduction, and equitable performance across groups. These are research priorities motivated by the mapped gap, not outcomes established by the bibliometric analysis itself.

Generative artificial intelligence makes the distinction between risk and response especially complex because the same technology can function as an intervention, relationship, information source, and exposure. Systematic review evidence identifies privacy, information integrity, fairness, user safety, and governance as major ethical domains for generative AI in mental health (Wang et al. 2025), while youth-facing evaluations raise concerns about risk monitoring and crisis handling (Clark 2025, Sobowale et al. 2025). The manual bridge validation in the present study is informative in this regard: several algorithmically identified Both records were not true direct bridges because AI was primarily the exposure or because the risk term was background context. This distinction strengthens the conceptual meaning of the integration gap and illustrates why simple keyword co-occurrence should not be equated with an integrated prevention pathway.

The timing of validated bridge records warrants cautious interpretation. All five manually validated direct bridges in complete years appeared in 2024-2025, whereas four more were present in the provisional 2026 snapshot. This concentration is compatible with an emerging line of inquiry linking cyberbullying detection, online-safety tools, AI-supported education, and precision prevention, but the complete-year number is too small to support a stable trend claim. Accordingly, 2026 was excluded from trend inference and shown only as provisional. The more defensible conclusion is that direct integration has become visible but remains rare.

The collaboration and citation networks add a second structural signal. Country citation mapping was highly centralized around the United States, while author co-authorship was more fragmented and cited-author co-citation contained several intellectual communities. The United States remained the leading country under both full and fractional publication and citation allocation, indicating that this geographic conclusion was not an artifact of counting method. Nevertheless, database choice matters: Scopus and other bibliographic databases differ in journal and disciplinary coverage (Singh et al. 2021), so geographic ranks should be read as properties of this Scopus-defined corpus rather than universal rankings. The uneven collaboration structure also raises questions about generalizability because regulatory environments, platform ecosystems, language, and access to adolescent mental-health services differ across settings.

For child and adolescent psychiatry, school health, and preventive mental-health research, the mapped structure suggests several testable directions. Future empirical studies could evaluate AI tools within complete prevention and referral pathways rather than as stand-alone accuracy products; high-stakes applications involving suicidality or self-harm could explicitly assess accountable human oversight; school-based cyberbullying prevention could be linked to early identification and referral mechanisms; platform-level outcomes such as exposure reduction, moderation quality, recurrence, and transparency could be studied alongside individual symptoms; and adolescents and families could be involved in co-design and safety evaluation. These recommendations are not direct bibliometric findings. They are

implications generated by combining the observed integration gap with the external clinical, implementation, and policy literature (Sharma et al. 2025, Wang et al. 2025, World Health Organization Regional Office for Europe 2025).

This study has several strengths. It used a large multidisciplinary corpus, transparent and reproducible corpus-definition rules, complete-year anchoring of temporal analyses, manual validation of the integrated-record set, record-level audits, complementary sensitivity analyses, and multiple science-mapping approaches. Thematic evolution, keyword burst detection, network-community validation, and alternative country-counting approaches further strengthened the robustness of the findings.

Several limitations should nevertheless be considered. First, Scopus was used as the sole bibliographic database; therefore, publications indexed exclusively elsewhere were not captured and citation indicators are database-specific. Second, retrieval and classification relied on explicit terms in bibliographic metadata, making the corpus sensitive to terminology and indexing practices despite the record-level audit and sensitivity analyses. Third, the operationalized integration gap reflects explicit co-framing of digital risk and AI-enabled response within the same publication and should not be interpreted as evidence of absent functional, clinical, or interdisciplinary integration across the wider research ecosystem. Fourth, the risk literature was strongly dominated by cyberbullying-related terminology, which limits generalization to all forms of digital harm. Finally, 2026 represented an incomplete publication year and was therefore treated descriptively rather than used for temporal trend inference. Network and country-level findings may also vary according to analytical thresholds, normalization choices, database coverage, and counting methods. Accordingly, the findings should be interpreted as evidence of bibliographic structure and research trends rather than clinical effectiveness or causal relationships.

Conclusion

Adolescent digital mental-health research has shifted toward substantially greater AI-enabled detection, prediction, and support activity, with response-focused output exceeding risk-focused output in the complete 2025 calendar year. However, explicit within-publication bibliographic co-framing of digital risk and AI-enabled response remains exceptionally uncommon: fewer than 1% of records met the deterministic Both definition across alternative corpus specifications, and manual validation produced an even smaller direct-bridge estimate. This finding identifies a bibliographic integration gap within individual publications; it should not be read as evidence that functional, clinical, interdisciplinary, or cross-publication integration is absent from the broader research ecosystem. Nor does the bibliometric analysis establish clinical effectiveness, prevention impact, or causal benefit. The next empirical step is to test how AI-enabled tools can be connected safely and equitably with developmentally appropriate prevention, referral, care, and governance systems.

References

Aboujaoude E, Savage MW, Starcevic V, Salame WO (2015) Cyberbullying: Review of an old problem gone viral. J Adolesc Health, 57:10-18.

Armitage R (2021) Bullying in children: Impact on child health. BMJ Paediatr Open, 5:e000939.

Baas J, Schotten M, Plume AM, Côté G, Karimi R (2020) Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. Quant Sci Stud, 1:377-386.

Babar A, Asif MS, Kharub I, Beydoun G (2026) Cyberbullying among adolescents: Mapping 25 years of global research. J Manag Organ, 32:1145-1167.

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp, 2008:P10008.

Boer M, Stevens GWJM, Finkenauer C, de Looze ME, van den Eijnden RJJM (2021) Social media use intensity, social media use problems, and mental health among adolescents: Investigating directionality and mediating processes. Comput Human Behav, 116:106645.

Bottino SMB, Bottino CMC, Regina CG, Correia AVL, Ribeiro WS (2015) Cyberbullying and adolescent mental health: Systematic review. Cad Saude Publica, 31:463-475.

Clark A (2025) The ability of AI therapy bots to set limits with distressed adolescents: Simulation-based comparison study. JMIR Ment Health, 12:e78414.

Dalouh Ounia R, Soriano Ayala E, Caballero Cala V (2025) Deepfakes and digital violence in adolescent couples: Perceptions, risks, and prevention strategies. Pedagogika, 160:74-93.

Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM (2021) How to conduct a bibliometric analysis: An overview and guidelines. J Bus Res, 133:285-296.

Dosovitsky G, Bunge E (2023) Development of a chatbot for depression: Adolescent perceptions and recommendations. Child Adolesc Ment Health, 28:124-127.

Feng Y, Hang Y, Wu W, Song X, Xiao X, Dong F et al. (2025) Effectiveness of AI-driven conversational agents in improving mental health among young people: Systematic review and meta-analysis. J Med Internet Res, 27:e69639.

Hamm MP, Newton AS, Chisholm A, Shulhan J, Milne A, Sundar P et al. (2015) Prevalence and effect of cyberbullying on children and young people: A scoping review of social media studies. JAMA Pediatr, 169:770-777.

John A, Glendenning AC, Marchant A, Montgomery P, Stewart A, Wood S et al. (2018) Self-harm, suicidal behaviours, and cyberbullying in children and young people: Systematic review. J Med Internet Res, 20:e129.

Kadirvelu B, Bellido Bel T, Freccero A, Di Simplicio M, Nicholls D, Faisal AA (2026) Digital phenotyping for adolescent mental health: Feasibility study using machine learning to predict mental health risk from active and passive smartphone data. J Med Internet Res, 28:e72501.

Kelly Y, Zilanawala A, Booker C, Sacker A (2018) Social media use and adolescent mental health: Findings from the UK Millennium Cohort Study. EClinicalMedicine, 6:59-68.

Kleinberg J (2002) Bursty and hierarchical structure in streams. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining:91-101. July 23-26, 2002, Edmonton, Alberta, Canada. New York, Association for Computing Machinery.

Montazeri A, Mohammadi S, Mokhtari-Hesari P, Ghaemi M, Riazi H, Sheikhi-Mobarakeh Z (2023) Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO): A minimum requirements. Syst Rev, 12:239.

O'Reilly M, Dogra N, Whiteman N, Hughes J, Eruyar S, Reilly P (2018) Is social media bad for mental health and wellbeing? Exploring the perspectives of adolescents. Clin Child Psychol Psychiatry, 23:601-613.

Pellegrino A, Stasi A (2024) A bibliometric analysis of the impact of media manipulation on adolescent mental health: Policy recommendations for algorithmic transparency. Online J Commun Media Technol, 14:e202453.

Radovic A, Gmelin T, Stein BD, Miller E (2017) Depressed adolescents' positive and negative use of social media. J Adolesc, 55:5-15.

Sharma G, Yaffe MJ, Ghadiri P, Gandhi R, Pinkham L, Gore G et al. (2025) Use of artificial intelligence in adolescents' mental health care: Systematic scoping review of current applications and future directions. JMIR Ment Health, 12:e70438.

Singh VK, Singh P, Karmakar M, Leta J, Mayr P (2021) The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. Scientometrics, 126:5113-5142.

Sobowale K, Humphrey DK, Zhao SY (2025) Evaluating generative AI psychotherapy chatbots used by youth: Cross-sectional study. JMIR Ment Health, 12:e79838.

Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P et al. (2021) The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry, 20:318-335.

van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84:523-538.

Viner RM, Aswathikutty-Gireesh A, Stiglic N, Hudson LD, Goddings AL, Ward JL et al. (2019) Roles of cyberbullying, sleep, and physical activity in mediating the effects of social media use on mental health and wellbeing among young people in England: A secondary analysis of longitudinal data. Lancet Child Adolesc Health, 3:685-696.

Wang X, Zhou Y, Zhou G (2025) The application and ethical implication of generative AI in mental health: Systematic review. JMIR Ment Health, 12:e70610.

WHO (2025) Mental health of adolescents. https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health (Accessed 15.08.2026).

World Health Organization Regional Office for Europe (2025) Addressing the Digital Determinants of Youth Mental Health and Well-Being: Policy Brief. Copenhagen, World Health Organization Regional Office for Europe.

Authors Contributions: The author(s) have declared that they have made a significant scientific contribution to the study and have assisted in the preparation or revision of the manuscript
Peer-review: Externally peer-reviewed.
Ethical Approval: This study relies exclusively on secondary data retrieved from public academic literature databases (Scopus). No human subjects or animals were involved; therefore, formal institutional review board approval, ethics committee clearance were not required.
Conflict of Interest: No conflict of interest was declared.
Financial Disclosure: No financial support was declared for this study.
Note: All search queries, raw dataset files, and visualization parameters used in this bibliometric analysis and listed as supplementary files in the text have been made publicly available on the Zenodo repository to ensure reproducibility. All supplementary tables, figures, and queries mentioned in the text are available at https://doi.org/10.5281/zenodo.23296739.