Review

Data-Driven Methods Supporting Mental Health and Well-Being Service Delivery: Umbrella Review

Salman Ahmad[1], PhD; Raymond Bond[1], PhD; Zeeshan Tariq[1], PhD; Gillian Cameron[1], MSc; Maurice Mulvenna[1], PhD; Edel Ennis[2], PhD; Kristel Miller[3], PhD; Daniel Buckley[1], BA

[1]School of Computing, Faculty of Computing, Engineering and the Built Environment, University of Ulster, Belfast, Northern Ireland, United Kingdom

[2]School of Psychology, Faculty of Life and Health Sciences, University of Ulster, Coleraine, Northern Ireland, United Kingdom

[3]Department of Management, Leadership & Marketing, Ulster University Business School, University of Ulster, Belfast, Northern Ireland, United Kingdom

Corresponding Author:

Salman Ahmad, PhD
School of Computing
Faculty of Computing, Engineering and the Built Environment, University of Ulster
2-24 York Street
Belfast, Northern Ireland BT15 1AP
United Kingdom
Phone: +44 28 9536 5040
Email: s.ahmad1@ulster.ac.uk

Abstract

Background: Mental health and well-being services generate substantial operational and client-level data. Review-level evidence on how data-driven methods support service delivery remains fragmented across technologies and settings.

Objective: This study aimed to synthesize review-level evidence on data-driven methods relevant to mental health and well-being service delivery, including business intelligence (BI), data visualization, and machine learning (ML), and identify applications, implementation challenges, and evidence gaps.

Methods: We conducted an umbrella review following Joanna Briggs Institute methodology and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020. The original searches of Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and PsycINFO were completed on January 27, 2026. During revision, supplementary PubMed and Embase searches were conducted on September 22, 2026. For both supplementary databases, the search was restricted at the database-search stage to publications from 2021 to the current date and to review/systematic review records; eligibility screening then retained the original January 27, 2026, evidence cutoff. All records were independently screened by 2 reviewers using Covidence (Veritas Health Innovation); the reviewer pair differed between the original and supplementary search phases. Eight structured reviews were critically appraised using the JBI checklist and synthesized narratively.

Results: Eight structured reviews were included (3 systematic reviews, 4 scoping reviews, and 1 umbrella review). Seven reviews contributed primary study or source-level evidence, while one umbrella review synthesized 44 reviews. The retrieved evidence was dominated by AI and ML, including conversational agents, natural language processing, and clinical decision support, whereas dedicated BI and operational data visualization syntheses were sparse. Two reviews directly addressed mental health service delivery, 4 were adjacent or partially direct, and 2 were indirect but relevant. Only one review focused specifically on mental health helplines. Service-level benefits were reported more consistently than sustained clinical benefits, although the evidence remained heterogeneous and partly indirect. Safety and implementation concerns included symptom worsening, bias, hallucination, privacy, trust, integration barriers, and inequitable access.

Conclusions: Review-level evidence relevant to mental health and well-being services is uneven and concentrated on AI/ML. Because the search required all 3 concept blocks, BI- and visualization-focused reviews may have been under-retrieved. BI and operational data visualization, helpline-specific implementation, long-term effectiveness, safety, and equity require more direct and prospective evaluation before findings can be generalized across service settings.

Trial Registration: OSF Registries 10.17605/OSF.IO/8NQBC; https://doi.org/10.17605/OSF.IO/8NQBC

J Med Internet Res 2026;28:e106104; doi: 10.2196/106104

Keywords: umbrella review; business intelligence; data visualisation; machine learning; mental health services; clinical decision support; conversational agents; natural language processing; mental health service delivery

Introduction

Mental health and well-being services, including counseling and psychotherapy, helplines, workplace employee assistance programs (EAPs), student support, veteran support, and digital mental health platforms, are under sustained pressure from rising demand, workforce constraints, and the long tail of population mental health needs intensified by the COVID-19 pandemic [1,2]. These services generate substantial operational and client-level data, ranging from referral and caseload records to encounter notes, helpline call logs, app interaction events, and free-text feedback. Translating these data into actionable insights is widely seen as a route to improved service planning, demand management, risk detection, engagement monitoring, and outcome evaluation [3].

Three broad classes of data-driven methods are most commonly proposed for this purpose. Business intelligence (BI) systems and dashboards consolidate operational data into views that support strategic and operational decision-making by service planners and managers [4]. Data visualization, including visual analytics and operational dashboards, supports interpretation of high-volume and multidimensional service data. Machine learning (ML) and broader AI methods, including natural language processing (NLP), conversational agents, predictive modeling, and deep learning, are increasingly applied to risk prediction, screening, conversational service delivery, and the analysis of unstructured patient feedback [5-7].

Research and review activity in this space has grown rapidly since 2020 and remains fragmented across disciplines, populations, and technology categories. Recent reviews have examined ML for suicide prediction [8], digital mental health platforms [9], digital mental health services during the COVID-19 pandemic [10], AI-based conversational agents in health care [11], digital therapeutics analytics [12], and AI and digital health in primary care in low- and middle-income countries (LMICs) [13]. These reviews vary in scope, methodology, and quality. Some are preregistered systematic reviews with formal critical appraisal; others are nonsystematic narrative reviews with no defined search or inclusion criteria. Their findings have not been integrated.

Two distinct questions motivate this review, and they sit at different levels of the evidence base. The first is integrative: multiple recent reviews now address overlapping aspects of AI, ML, NLP, and conversational agents in mental health and well-being service contexts, yet they differ in rigor, scope, and disciplinary framing, and their findings have not been brought together. An umbrella review is the appropriate design for this question precisely because the unit of analysis is the review itself: the task is to appraise and reconcile a set of existing syntheses, not to resynthesize primary studies [14,15]. The second question is diagnostic: it is not known whether BI and operational data visualization, which are routine in mental health service organizations, have been synthesized at the review level at all. Mapping the review-level literature can establish how these technologies are represented within retrieved evidence syntheses. A central contribution of the present review is therefore to characterize the relative sparsity of BI and operational data visualization within the retrieved review-level literature. Given the search architecture, this finding should not be interpreted as proof that no such reviews exist in the wider literature. An umbrella review is distinguished from a scoping review in that it synthesizes only existing reviews, formally appraises those reviews, and assesses overlap in primary studies across the included reviews [14]. It is distinguished from a systematic review of primary studies in that the unit of analysis is the review itself.

One of the reviews eligible for inclusion was itself an umbrella review. Huynh et al [11] synthesizes 44 reviews of AI-based conversational agents across health care. Its presence within the evidence base does not make the present review redundant; it sharpens the rationale for it. The existing umbrella review was bounded to a single technology class, conversational agents, within health care at large, with mental health as one application domain among many and clinical health-related outcomes as its end point. The present umbrella review is bounded instead by a service context: mental health and well-being service delivery. It also spans a broader family of data-driven methods, namely BI, operational data visualization, predictive analytics, and ML, of which conversational agents are only one part, with service delivery and decision-making outcomes rather than clinical outcomes alone as their focus. The 2 reviews therefore answer different questions, and the existing umbrella review is incorporated as one input to the higher-order synthesis presented here rather than duplicated by it.

The present umbrella review synthesizes review-level evidence on the use of BI, data visualization, and ML in mental health and well-being services, focused on service delivery and decision-making rather than clinical diagnostic AI. The function of this review is to raise awareness among service providers of the potential uses of routinely collected data to enhance service provision and inform decisions about adopting into routine practice where such approaches may not currently be considered. The review is organized around 5 research questions (RQs):

- RQ1. What data-driven methods, tools, algorithms, and datasets are reported across the included reviews?
- RQ2. In which mental health service contexts are these methods applied (eg, counseling, EAPs, helplines, emergency departments, and digital platforms)?
- RQ3. What applications are reported, including risk detection, client segmentation, engagement monitoring, and service planning?

- RQ4. What benefits, insights, challenges, and barriers to implementation are reported?
- RQ5. What research gaps and directions for future inquiry can be identified? (This is addressed in the Discussion section.)

Methods

Protocol and Registration

This umbrella review followed the Joanna Briggs Institute (JBI) methodology for umbrella reviews [14] and is reported in line with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement [16]. A time-stamped Open Science Framework (OSF) registration draft was created on December 2, 2025, initially as a scoping review protocol. The final database searches were completed on January 27, 2026, using the search strategy developed for that scoping review protocol. On May 6, 2026, the protocol was revised to an umbrella review design and formally registered on OSF [17] because the available literature consisted predominantly of existing reviews, making higher-order synthesis the more informative and methodologically appropriate approach. The original database search strategy was retained; restriction to review document types was applied during screening rather than at the database-query stage. The design change therefore altered the unit of analysis and formal eligibility to review-level evidence, and umbrella review procedures for critical appraisal and assessment of primary-study overlap were applied. Formal screening reported in this manuscript was conducted within the dedicated Covidence (Veritas Health Innovation) review for the umbrella review. Because formal umbrella review registration occurred after the final database searches, it is not presented as prospective registration of the umbrella review search stage. The completed PRISMA 2020 checklist is provided in Checklist 1.

Eligibility Criteria

Criteria Overview

Eligibility criteria followed the JBI population-concept-context (PCC) structure adapted for umbrella reviews, where the unit of analysis is an existing review.

Inclusion Criteria

The following inclusion criteria were applied:

- Type of evidence: the formal umbrella review evidence base was restricted to structured evidence syntheses, comprising systematic reviews, scoping reviews, meta-analyses, and umbrella reviews that reported a defined search strategy, explicit eligibility criteria, and a structured synthesis method. Narrative reviews, conceptual reviews, mini-reviews, commentaries, and other nonsystematic publications were excluded from the formal synthesis and could be cited only as context in the Introduction or Discussion.
- Population: users, clients, or service-recipients of mental health or well-being services in nonclinical or service-delivery contexts (eg, counseling, EAPs, helplines, student support, workplace well-being, veteran support, and digital mental health platforms).
- Concept: application of data analytics, BI, dashboards, data visualization, ML, or predictive modeling.
- Context: use of these methods to support service delivery, decision-making, client experience, outcome monitoring, or service planning.
- Publication characteristics: peer-reviewed journal article or conference paper; English language; published between 2021 and 2026.

Exclusion Criteria

Primary studies; narrative reviews, conceptual reviews, mini-reviews, commentaries, and other nonsystematic publications that lacked a defined search strategy, explicit eligibility criteria, and a structured synthesis method; reviews focused exclusively on clinical diagnostic AI applications (eg, magnetic resonance imaging, electroencephalography, and neuroimaging-based diagnostic models) in which the application was not embedded in service-delivery decision-making; reviews unrelated to data analytics, BI, data visualization, or ML; non-English texts; and preprints and reviews not subjected to peer review.

In line with JBI umbrella review methodology, no structured review that met the eligibility criteria was excluded solely on the basis of its critical appraisal findings. Appraisal was used to inform interpretation rather than to gate inclusion. Two nonsystematic publications included in the originally submitted version of this manuscript (Lee et al [12] and Mumtaz et al [18]) were therefore removed from the formal evidence base during revision; when cited, they are identified as contextual literature only.

Information Sources and Search Strategy

Five electronic bibliographic databases were searched to capture relevant review-level literature across health, psychology, and computing disciplines: Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and PsycINFO. Searches were limited to peer-reviewed, English-language literature published between 2021 and January 27, 2026. Reference lists of included reviews were hand-searched for additional reviews. During manuscript revision, supplementary searches were conducted in PubMed and Embase on September 22, 2026, in response to reviewer concerns about biomedical database coverage.

The original database set was selected to reflect the interdisciplinary scope of the review rather than a single clinical discipline. Scopus and Web of Science provided broad multidisciplinary coverage, PsycINFO provided focused coverage of psychology and mental health, and IEEE Xplore and ACM Digital Library were included to capture computing and informatics literature. CINAHL and the Cochrane Library were not searched directly. The supplementary PubMed and Embase searches improved biomedical coverage and were not treated as a rerun of the original search.

The original search strategy used three mandatory concept blocks combined with AND: (1) mental health and well-being services; (2) broader data-driven methods, including data analytics, AI, ML, NLP, predictive modeling, data mining, clustering, and classification; and (3) BI, data visualization, and service or decision-making terminology. Full strategies for the original databases and the supplementary PubMed and Embase searches are provided in Multimedia Appendix 1; the PsycINFO strategy additionally combined APA Thesaurus controlled vocabulary with free-text terms (Multimedia Appendix 2). In both supplementary databases, search-stage restrictions were applied for publication date (2021 to September 22, 2026) and review/systematic review records. Because these searches ran after the original cutoff, the January 27, 2026, evidence cutoff was then enforced during eligibility screening; post-cutoff records were not incorporated into the formal synthesis. Restriction to review-type publications was applied during screening for the original 5 database search rather than at the query level. During development of the original scoping review strategy, retrieval was checked against 5 known relevant primary-study papers; this was not a formal validation set for review-level BI, dashboard, or data visualization evidence. Because all 3 Boolean concept blocks were mandatory, a review focused on BI, dashboard, or data visualization but lacking broader data-driven method terminology could still have been missed.

The original searches retrieved 1349 records: Scopus 545, Web of Science 375, PsycINFO 66, IEEE Xplore 250, and ACM Digital Library 113. The supplementary searches retrieved 73 additional records: PubMed 30 and Embase 43. Across both phases, 1422 database records were therefore identified. Covidence identified 304 duplicates in the original import, and 2 additional duplicates were removed manually; the supplementary imports contributed 40 further duplicates (15 from PubMed and 25 from Embase), for a total of 346 duplicates removed overall.

Study Selection

All retrieved records were managed in Covidence (Veritas Health Innovation). For the original 5-database search, 2 reviewers (SA and EE) independently screened title/abstract records and full-text reports against the predefined eligibility criteria, and DB contributed to title/abstract screening; conflicts were discussed in review team meetings involving GC and MM, with MM making the final eligibility decision where consensus was not reached. The original database queries were not restricted to review document types, so nonreview records were removed during screening, and full-text exclusions were recorded with reasons. For the revision-stage PubMed and Embase searches, the 33 deduplicated supplementary records were independently screened in Covidence by SA and GC using the same eligibility criteria and the evidence cutoff. Both reviewers assessed the 2 reports that progressed to full text, and both met the eligibility criteria; Wah [19] and Hazhir et al [20] were therefore incorporated. Hazhir et al [20] was newly identified by the supplementary search. A record for Wah [19] had also been retrieved by the original search and excluded at original title and abstract screening because its title and abstract presented a broad eHealth and health care–chatbot scope, which the screeners judged to be outside mental health service delivery; full-text review after the supplementary search showed that mental health support was an explicit application domain. The supplementary record was not flagged by Covidence as a duplicate of that record. The same eligibility criteria and evidence cutoff were applied to the original and supplementary records; only the reviewer pair differed.

Critical Appraisal of Included Reviews

Each included structured review was appraised using the 11-item JBI critical appraisal checklist for systematic reviews and research syntheses. This single checklist was applied across systematic, scoping, and umbrella reviews because it is the instrument specified in JBI umbrella review methodology for research syntheses [14], and a common instrument allowed item-by-item comparisons across the evidence base; items that did not apply to a review's design were recorded as not applicable (NA). Appraisal was conducted independently by 2 reviewers (SA and GC), with disagreements resolved through discussion and, where necessary, adjudication by a third reviewer (EE). Across the 8 formally included reviews, the 2 independent reviewers agreed on 70 of 88 item-level judgments before adjudication (79.5% raw agreement); the 18 disagreements were resolved through the established adjudication process. Responses were recorded as "Yes," "No," "Unclear," or "NA." NA was retained as a distinct applicability judgment rather than treated as a methodological failure or converted to 0. Item 5 (appraisal criteria) was judged NA where a review undertook no critical appraisal of its included studies, and item 6 (independent appraisal by 2 or more reviewers) was judged "No" where independent dual appraisal was not reported; at final revision, 2 adjudicated judgments for the reviews added at revision (Wah [19], item 5; Hazhir et al [20], item 6) were harmonized with this convention. Because applicability differed across scoping, systematic, and umbrella review designs, item responses were not converted into composite numerical scores or high/moderate/low quality categories. Critical appraisal was interpreted descriptively at item level and used to temper conclusions; no review meeting the eligibility criteria was excluded solely on appraisal findings.

The final adjudicated item-level judgments for all 8 structured reviews are reported in the Results. Lee et al [12] and Mumtaz et al [18] are nonsystematic contextual publications and were not included in the formal JBI appraisal.

Assessment of Primary-Study Overlap and Corrected Covered Area

Because included reviews can share underlying evidence, we constructed a citation matrix and calculated the corrected covered area (CCA) using the formula established by Pieper et al [21]: CCA=(N-r)/(r×(C-1)), where N is the total number of included-unit occurrences, r is the number of unique units after deduplication, and C is the number of reviews evaluated in the tier. A single omnibus CCA was not appropriate because 7 included reviews synthesized primary study

or source-level evidence, whereas Huynh et al [11] was an umbrella review of 44 reviews. We therefore calculated CCA at 2 primary-study tiers: (1) the 7 primary-level reviews (C=7; N=264 included study/source occurrences); and (2) the direct mental health service-delivery subset comprising Dwyer et al [22] and Gu et al [23] (C=2; N=91). Unique units (r) were identified from the included studies or sources reported by each review. Where part of a review's included-study list could not be identified from the published article, we also calculated a worst-case CCA that assumed that every unidentified study duplicated a study in another review. In Tier 3, overlap between the 44 constituent reviews of Huynh et al [11] and the 7 standalone reviews was assessed by direct counting rather than by CCA because the comparison set at that level was the list of reviews included in the present umbrella review, which was not itself a constituent review. Overlap matching used author, title, publication year, and DOI where available; topic similarity, chatbot names, or shared surnames alone were not treated as matches. CCA values were interpreted using the thresholds described by Pieper et al [21].

Cross-tier relationships were reported separately rather than counted as within-tier duplication because they represent different units of analysis. Full calculations, citation matrices, and limitations are reported with the Results.

Data Extraction

Data were initially extracted from each included review by the lead reviewer (SA) using a structured data-extraction form (Multimedia Appendix 3). The form captured bibliographic details; review type and stated methodology; review aims and RQs; number and type of primary studies included; population and service context; data-driven methods; tools, platforms, datasets, and algorithms; reported applications, outcomes, benefits, challenges, and limitations; authors' conclusions and stated evidence gaps; and item-level JBI critical appraisal judgments. A second reviewer (GC) independently cross-checked the completed extraction against the source reviews. Differences or ambiguities identified during this verification were discussed by SA and GC and resolved by consensus; where source reporting remained unclear, the extracted entry was qualified rather than inferred. When an included review reported an individual-study effect estimate, diagnostic-performance figure, or named study example, this information was treated as secondary-reported data from that review rather than as an independently assessed primary-study finding. Primary studies were not independently retrieved, critically appraised, or reanalyzed to verify these figures. A provenance table links such examples to the included reviews from which they were extracted.

Synthesis

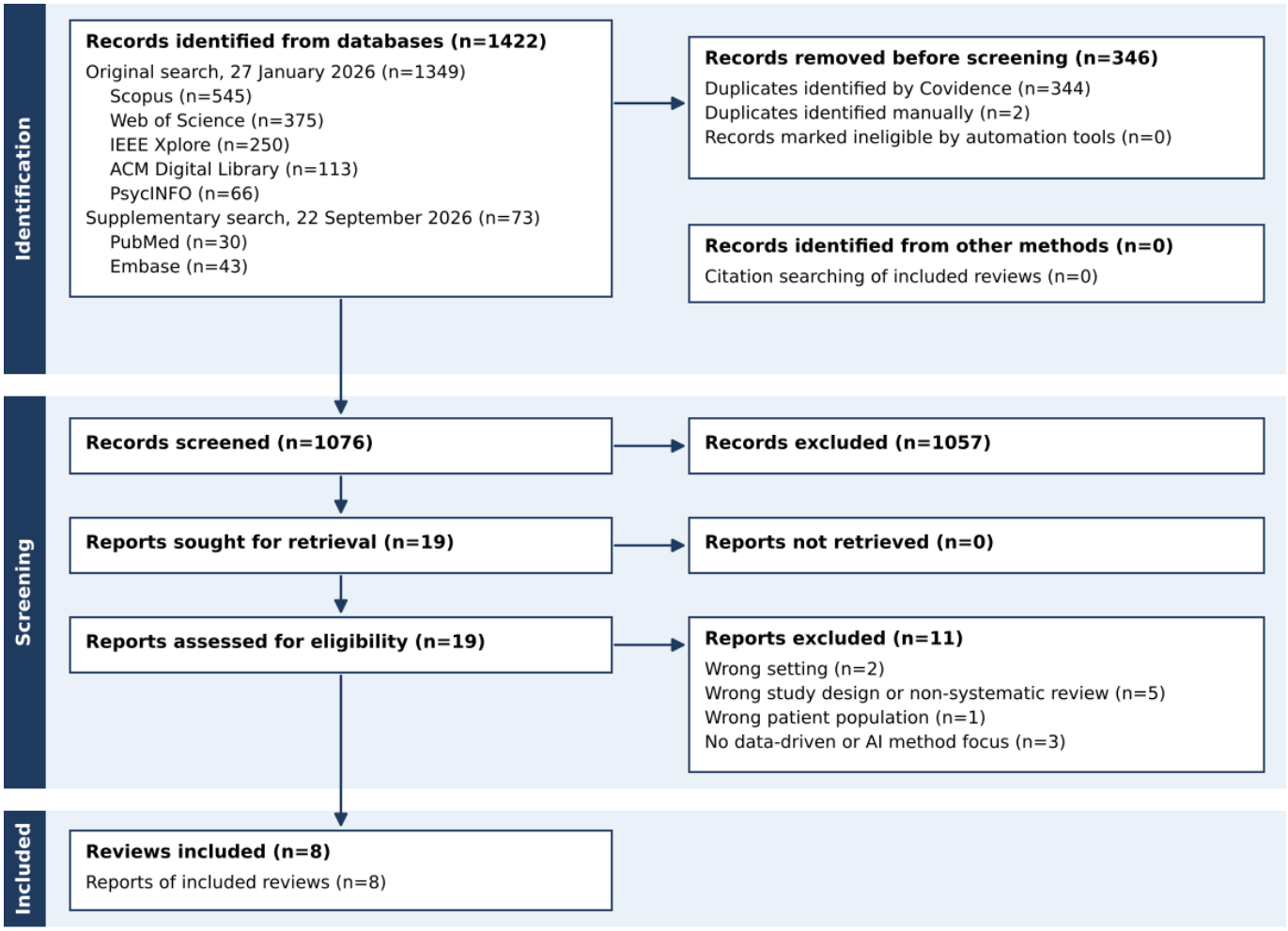
Quantitative meta-analytic resynthesis was not planned because of heterogeneity in review scope, methods, and outcomes. Findings were synthesized narratively in line with JBI umbrella review guidance and structured around 5 RQs. RQ1-RQ3 were summarized in a concise main-text matrix, supported by a full review-by-review matrix of methods, service contexts, applications, and evidence directness. RQ4 was synthesized separately, and RQ5 was addressed in the Discussion. Service-level findings and technology-level findings were distinguished explicitly. To make applicability to mental health service delivery explicit, each formally included review was classified during revision as direct, adjacent, partially direct, or indirect yet relevant evidence. SA assigned the classifications, and GC verified them against the source reviews; differences were resolved by discussion. This reporting classification did not alter eligibility or appraisal and was used to calibrate interpretation and transferability across service contexts.

Results

Search Results and Study Selection

Across the original and supplementary searches, 1422 database records were identified: 1349 from the original 5 databases and 73 from the supplementary PubMed and Embase searches. A total of 346 duplicates were removed (306 from the original search workflow and 40 from the supplementary imports), leaving 1076 records for title and abstract screening. Of these, 1057 were excluded at the title and abstract stage, including 31 of the 33 supplementary records. Nineteen reports were sought and assessed at full-text review (17 from the original search and 2 from the supplementary searches). Eleven full-text reports were excluded, all from the original search, for the following reasons: wrong setting (n=2), wrong study design or nonsystematic review (n=5), wrong patient population (n=1), and no data-driven or AI method focus (n=3). Both supplementary full-text reports met the eligibility criteria, resulting in 8 structured reviews in the final formal synthesis (Figure 1). Hand-searching of the reference lists of included reviews identified no additional reviews.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram. The final flow integrates the original 5-database search with the revision-stage PubMed and Embase searches while preserving the original January 27, 2026, evidence cutoff.



Supplementary revision-stage searches (PubMed and Embase, 22 September 2026): 33 records remained after deduplication and were screened independently by SA and GC; 31 were excluded at title/abstract, and 2 reports were assessed at full text by both reviewers and included. Eligibility retained the original 27 January 2026 evidence cutoff.

Characteristics of Included Reviews

Table 1 summarizes the characteristics of the 8 structured reviews included in the formal umbrella synthesis. The set comprises 3 systematic reviews [13,19,23], 4 scoping reviews [20,22,24,25], and 1 umbrella review [11]. Six reviews reported 235 included studies in total (Chew and Achananuparp [24], 26; Dwyer et al [22], 70; Feizollah et al [25], 52; Gu et al [23], 21; Saif-Ur-Rahman et al [13], 48; and Hazhir et al [20], 18), predominantly primary studies; Wah [19] reported 29 included sources of mixed types rather than a wholly primary-study set, and Huynh et al [11] synthesized 44 reviews. Lee et al [12] and Mumtaz et al [18] remain outside the formal evidence base and are cited only contextually where relevant.

Table 1. Characteristics of included structured reviews (n=8).

Number	First author (year)	Review type	Search dates	Databases	Included studies or sources (n)	Conditions/scope	Geographic scope	Methodological framework
1	Chew and Achananuparp (2022) [24]	Scoping	From inception to June 21, 2021	ACM Library, CINAHL, Cochrane Central, Embase, IEEE Xplore, PsycINFO, PubMed, Scopus, and Web of Science	26	Health care broadly; chronic disease self-management; 2 MH[a]-specific	Mixed (11 countries)	Arksey and O'Malley
2	Dwyer et al (2021) [22]	Scoping	From inception to November 24, 2020	PsycINFO, PubMed, ERIC[b], and Web of Science	70	Text-based MH communications; rural/remote framing	Predominantly urban/suburban HIC[c]	PRISMA-ScR[d]

Number	First author (year)	Review type	Search dates	Databases	Included studies or sources (n)	Conditions/scope	Geographic scope	Methodological framework
3	Feizollah et al (2025) [25]	Scoping (JBI[e])	February 2024 (no date limit)	MEDLINE, Embase, CINAHL, Cochrane Database of Reviews, and Google Scholar	52	NLP[f] of unstructured patient feedback across health care	Mixed; English- and non-English-speaking nations	JBI scoping methodology
4	Gu et al (2024) [23]	Systematic	April 19, 2022; updated September 2023	Scopus, PubMed, PsycINFO	21	CDSS[g] in mental health helpline services; included studies extended beyond the stated helpline focus to telepsychiatry and primary care	Predominantly HIC	PRISMA[h] 2020
5	Huynh et al (2026) [11]	Umbrella	January 1, 2000, to August 20, 2025	MEDLINE, APA PsycInfo, CINAHL, Cochrane DSR[i], Academic Search Complete, Web of Science	44 reviews	AI conversational agents in health care; 23/44 reviews MH-focused	Mixed	JBI umbrella; PRISMA
6	Saif-Ur-Rahman et al (2023) [13]	Systematic	Oct 16, 2020; updated June 29, 2023	MEDLINE/PubMed, Web of Science, Embase, Cochrane Library	48	AI and digital health in primary care service delivery in LMICs[j]; 6/48 studies MH-focused	LMIC, 21 countries	PRISMA 2020; PROSPERO[k] registered
7	Wah (2025) [19]	Systematic	Search date not clearly reported; included sources prioritized 2022-2025	PubMed, IEEE Xplore, Web of Science; 4 additional records from unspecified other sources	29 included sources[l]	Broad health care/eHealth; MH support explicit; MH-specific source count not reported	International/ mixed	Systematic approach; PRISMA
8	Hazhir et al (2026) [20]	Scoping	From inception to February 18, 2025	PubMed, Scopus, Web of Science, Embase, ProQuest, and EBSCO[m]	18	Health care access for underserved/vulnerable populations; 6/18 studies MH-specific	10 countries; United States n=5	Arksey and O'Malley; PRISMA-ScR

[a]MH: mental health.
[b]ERIC: Education Resources Information Center.
[c]HIC: high-income country.
[d]PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews.
[e]JBI: Joanna Briggs Institute.
[f]NLP: natural language processing.
[g]CDSS: clinical decision support system.
[h]PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
[i]DSR: Database of Systematic Reviews.
[j]LMIC: low- and middle-income country.
[k]PROSPERO: International Prospective Register of Systematic Reviews.
[l]Sources of mixed type, including secondary reviews, book chapters, preprints, and web pages.
[m]EBSCO: Elton B Stephens Company.

Geographic and population scopes varied considerably. Dwyer et al [22] and Gu et al [23] provided the 2 direct review-level sources for mental health service delivery. Huynh et al [11], Saif-Ur-Rahman et al [13], Wah [19], and Hazhir et al [20] were classified as adjacent or partially direct because each included a substantive or explicit mental health component within a broader health care or primary care scope. Chew and Achananuparp [24] and Feizollah et al [25] were classified as indirect yet relevant. Gu et al [23] was the only included review explicitly focused on clinical decision support systems (CDSSs) in mental health helpline services.

All 8 formally included publications were structured evidence syntheses meeting the umbrella review eligibility criteria. Lee et al [12] and Mumtaz et al [18] were cited only as contextual literature and were not included in this table. Rows follow the review numbering (P1-P8) used throughout. Although Huynh et al [11] and Hazhir et al [20] have 2026 volume dates, they were published online on November 26, 2025, and December 30, 2025, respectively, before the January 27, 2026, evidence cutoff.

Methodological Quality

Item-level JBI critical appraisal findings for the 8 structured reviews are presented in Table 2 and Multimedia Appendix 4. All 8 reviews clearly stated a review question or objective, provided appropriate inclusion criteria, and all provided recommendations and future research directions supported by their syntheses. The search strategy was judged yes for 7 reviews and unclear for Wah [19]. The adequacy of information sources was judged yes for 6 reviews and unclear for Gu et al [23] and Wah [19]. Appraisal criteria were judged yes for Huynh et al [11] and Saif-Ur-Rahman et al [13] and NA for the other 6 reviews, none of which undertook formal critical appraisal of included studies. Independent critical appraisal by at least 2 reviewers was judged yes for Huynh et al [11] and Saif-Ur-Rahman et al [13] and no for the other 6 reviews. Methods to minimize extraction errors were judged yes for 6 reviews and unclear for Chew and Achananuparp [24] and Gu et al [23]; methods used to combine studies were judged yes for 5 reviews and unclear for Chew and Achananuparp [24], Gu et al [23], and Wah [19]. Publication

bias assessment was generally absent, unclear, or NA. These item-level limitations informed the interpretation.

Table 2. JBI[a] critical appraisal item-level judgments (final, following third-reviewer [EE] adjudication).

JBI item	P1 [24]	P2 [22]	P3 [25]	P4 [23]	P5 [11]	P6 [13]	P7 [19]	P8 [20]
Review question clearly stated	Y[b]	Y	Y	Y	Y	Y	Y	Y
Inclusion criteria appropriate	Y	Y	Y	Y	Y	Y	Y	Y
Search strategy appropriate	Y	Y	Y	Y	Y	Y	U[c]	Y
Sources and resources adequate	Y	Y	Y	U	Y	Y	U	Y
Appraisal criteria appropriate	[d]—	—	—	—	Y	Y	—	—
Critical appraisal by ≥2 reviewers	N[e]	N	N	N	Y	Y	N	N
Methods to minimize extraction errors	U	Y	Y	U	Y	Y	Y	Y
Methods to combine studies appropriate	U	Y	Y	U	Y	Y	U	Y
Publication bias assessed	N	N	N	N	U	—	N	N
Recommendations supported by data	Y	Y	Y	Y	Y	Y	Y	Y
Future research directives appropriate	Y	Y	Y	Y	Y	Y	Y	Y

[a]JBI: Joanna Briggs Institute.
[b]Y: Yes.
[c]U: Unclear.
[d]Not applicable. It indicates that an item was judged not applicable to the methodology or synthesis undertaken and was not treated as an adverse quality judgment or converted to zero. Appraisal findings were interpreted at the item level; no composite numerical score or high/moderate/low quality band was assigned. Item 5 (appraisal criteria) was judged not applicable where a review undertook no critical appraisal of its included studies; item 6 (critical appraisal by ≥2 reviewers) was judged No (N) where independent dual appraisal was not reported. Final judgments reflect third-reviewer (EE) adjudication of conflicts between the 2 independent appraisers (SA and GC), with P7 item 5 and P8 item 6 harmonized with this convention at final revision.
[e]N: No.

Two review-specific methodological cautions are relevant. First, the umbrella review by Huynh et al [11] reported a per-item JBI appraisal of its 44 included reviews; however, the published appraisal table was not fully legible, limiting independent verification of those review-level quality claims. Second, Wah [19] reported a structured review process without providing a fully reproducible search strategy or sufficiently specifying the quantitative/statistical synthesis to which it referred; its numerical effectiveness claims were therefore treated as secondary-reported findings and interpreted cautiously (Multimedia Appendix 5). Wah [19] was nonetheless retained as a structured evidence synthesis because it named the databases searched (PubMed, IEEE Xplore, and Web of Science), states eligibility criteria, described its screening process and secondary-reviewer validation of extracted data, and reported a thematic synthesis. Its 29 included sources contained secondary reviews, book chapters, preprints, and webpages. Wah [19] was used only as adjacent evidence alongside other reviews and was not the sole basis for any conclusion of the present review. Hazhir et al [20] did not undertake formal critical appraisal of included primary studies, consistent with its scoping review design.

Summary of Methods, Service Contexts, Applications, and Evidence Directness

Table 3 provides a concise review-level summary of the dominant method family, service context, principal application, and evidence directness for each included review. The full detailed evidence matrix is provided in Multimedia Appendix 6.

Table 3. Concise review-level summary of methods, service contexts, applications, and evidence directness.

Number	Review	Dominant methods	Service context	Principal applications	Evidence directness
1	Chew and Achananuparp (2022) [24]	AI adoption, conversational agents, NLP[a], and predictive methods	General/community health care and 2/26 studies MH[b]-specific	Acceptability, trust, privacy, and adoption	Indirect yet relevant
2	Dwyer et al (2021) [22]	NLP, sentiment analysis, and ML[c]	Text-based MH counseling and helplines	Service delivery, language-based outcome/risk signals, and engagement	Direct
3	Feizollah et al (2025) [25]	NLP, sentiment analysis, topic modeling, and text classification	Health care patient feedback and limited MH-specific evidence	Feedback analytics and service-quality monitoring	Indirect yet relevant
4	Gu et al (2024) [23]	Clinical decision support systems	MH helplines and included studies also extended to adjacent settings	Risk assessment, monitoring, and clinician decision support	Direct, with scope qualification

Number	Review	Dominant methods	Service context	Principal applications	Evidence directness
5	Huynh et al (2026) [11]	AI conversational agents, NLP, and LLMs[d]	Health care broadly and 23/44 reviews MH-focused	Clinical decision support, MH intervention, and self-management	Adjacent or partially direct
6	Saif-Ur-Rahman et al (2023) [13]	Digital health/mHealth[e] and limited substantive ML	LMIC[f] primary care and 6/48 studies MH-specific	Service delivery, screening, decision support, and adherence	Adjacent or partially direct
7	Wah (2025) [19]	Hybrid AI chatbots, NLP, ML, deep learning, LLMs, EHR[g]/IoT[h] integration	Health care broadly, MH support explicit, and MH-specific source count not reported	Engagement, triage, follow-up, monitoring, MH support, and workflow efficiency	Adjacent or partially direct
8	Hazhir et al (2026) [20]	Rule-based chatbots (n=13), hybrid rule-based+NLP (n=4), NLP-only (n=1)	Underserved/vulnerable populations and 6/18 studies MH-specific	Access, counseling/education, distress screening, psychological first aid, grief support, and referral	Adjacent or partially direct

[a]NLP: natural language processing.
[b]MH: mental health.
[c]ML: machine learning.
[d] LLM: large language model.
[e]mHealth: mobile health.
[f]LMIC: low- and middle-income country.
[g]EHR: electronic health record.
[h]IoT: Internet of Things.

RQ1: Data-Driven Methods, Tools, Algorithms, and Datasets Used (BI, Visualization, ML, and Predictive Analytics)

Across the 8 structured reviews, the methods landscape was heavily weighted toward AI and ML approaches, with BI and operational data visualization thinly covered.

The largest method cluster was AI-based conversational agents and NLP. Huynh et al [11] synthesized 44 reviews of AI-based conversational agents in health care, with deep learning (n=14 reviews), NLP (n=12), natural language understanding (n=5), and natural language generation (n=2) being the dominant underlying techniques, alongside large language models (LLMs; eg, ChatGPT, Bard, Claude, and Bidirectional Encoder Representations from Transformers [BERT]) in recent reviews. Conversational agent types covered text-based agents (n=18 reviews), voice-based agents (n=15), embodied agents (n=10), and mixed text-and-voice agents (n=6). Dwyer et al [22] reviewed text-based e-mental health services and the use of computational linguistic tools to predict mental health status from language use. The dominant tool was linguistic inquiry and word count (LIWC), with supplementary use of sentiment analysis, syntactic complexity measures, and supervised ML (Naive Bayes and regression models) applied to social media and online forum corpora. Feizollah et al [25] reviewed NLP applied to unstructured patient feedback across health services. It identified 3 method categories: sentiment analysis (n=32 included studies; tools included Valence Aware Dictionary and Sentiment Reasoner (VADER), support vector machines (SVMs), Naive Bayes, and a small number using transformer-based neural networks), topic modeling (n=15; predominantly latent Dirichlet allocation [LDA] and variants), and text classification (n=7; supervised and unsupervised learning).

The second method cluster was CDSS. Gu et al [23] synthesized 21 primary studies on CDSS design and implementation in mental health helpline services. Gu et al [23] classified the reviewed CDSSs into passive (n=12 studies), active (n=7), and semiactive (n=2) systems, with architectures consistently combining a database or knowledge base, a reasoning engine, and a user interface accessed via web or app, sometimes integrated with electronic health records (EHRs) via APIs.

A third method cluster was health care chatbot design and delivery. Wah [19] reviewed AI-powered hybrid chatbots combining automated AI with human input or oversight, including NLP, ML, deep learning, generative AI and LLMs, and EHR and Internet of Things (IoT) integration. Hazhir et al [20] characterized 18 chatbots for underserved and vulnerable populations: 13 were rule-based, 4 combined rule-based logic with NLP, and 1 was NLP-only; all used text-based interaction, with social media and web delivery most common. These 2 reviews expanded the conversational-agent evidence while remaining broader than mental-health-service-only syntheses.

A further method cluster concerned digital health and mobile health (mHealth) in LMIC primary care. In Saif-Ur-Rahman et al [13], only 1 of 48 included studies used ML (for depression screening); most evaluated mobile apps, SMS text messaging, interactive voice response (IVR), EHR access via mobile, or other digital health tools. This finding was specific to that LMIC primary-care review and should not be generalized to the wider mental health evidence base.

Chew and Achananuparp [24] focused primarily on perceptions, adoption, trust, privacy, and acceptability of AI in health care and provided comparatively limited algorithm-level detail. It explicitly excluded “machine learning” as a search keyword because of excessive irrelevant retrieval, which constrained its contribution to methods mapping. Wah

[19] similarly covered broad health care chatbot implementation themes, while Hazhir et al [20] provided more explicit design-level counts for rule-based, hybrid, and NLP-only systems.

At the tool, platform, and algorithm level, reporting across the 8 structured reviews remained uneven. Feizollah et al [25] named VADER, SVMs, Naive Bayes, transformer-based models, and LDA; Dwyer et al [22] identified LIWC and Naive Bayes; and Huynh et al [11] cataloged deep learning, NLP, and LLMs underpinning conversational agents. Wah [19] discussed hybrid chatbot architectures, GPT/BERT-family models, EHR integration, and IoT-enabled monitoring without reporting a sufficiently reproducible quantitative synthesis. Hazhir et al [20] reported rule-based versus hybrid/NLP chatbot designs and delivery platforms. Saif-Ur-Rahman et al [13] reported only one ML implementation across 48 studies, while Gu et al [23] reported CDSS mainly at the architectural level. Dataset-level reporting remained sparse, limiting appraisal of provenance, representativeness, and reproducibility.

BI and data visualization as distinct technology categories were sparsely represented across the 8 structured reviews. No formally included review centrally addressed BI dashboards, visual analytics, or operational data visualization for mental health and well-being service delivery. Gu et al [23] addressed user-interface design within CDSS, including color-coded risk summaries and dashboard-like interfaces, without providing a standalone visualization synthesis. This pattern was treated as a finding within the retrieved review-level evidence rather than definitive evidence that dedicated BI or visualization reviews do not exist.

RQ2: Mental Health Service Contexts Addressed

Overview

The service contexts addressed across the included reviews spanned clinical, community, digital, primary care, and health care access settings, with substantial differences in directness to the review question. Two reviews were classified as direct evidence for mental health service delivery (Dwyer et al [22] and Gu et al [23]); 4 as adjacent or partially direct (Huynh et al [11], Saif-Ur-Rahman et al [13], Wah [19], and Hazhir et al [20]); and 2 as indirect yet relevant (Chew and Achananuparp [24] and Feizollah et al [25]). These classifications are shown in Table 3 and were used to avoid treating broader health care evidence as equally transferable to specific mental health service contexts.

Direct Evidence: Helpline and Tele-Counseling Services

Gu et al [23] was specifically focused on mental health helpline services and was therefore classified as direct evidence. However, its included primary studies extended beyond strictly defined helplines to encompass telepsychiatry for nonpsychiatrists, primary care–based digital decision support (eg, a digital screening and decision-support e-tool in the Netherlands), and online cognitive behavioral therapy for depression. Its directness was therefore qualified by this broader effective scope, and helpline-specific conclusions were interpreted cautiously.

Direct Evidence: Text-Based Mental Health Services

Dwyer et al [22] was classified as direct evidence because it reviewed text-based e-mental health counseling and communication services, including helplines and commercial counseling platforms. Although the review was framed around rural and remote populations as an underserved group, the evidence it synthesized was dominated by urban and suburban adolescent and adult populations. The review was therefore direct to mental health service delivery, although less direct to its stated rural/remote population framing.

Adjacent or Partially Direct Evidence: Primary Care and Health Care Access

Saif-Ur-Rahman et al [13] addressed digital health and AI across LMIC primary health care rather than mental health services specifically; 6 of 48 primary studies were mental-health-focused. Hazhir et al [20] similarly addressed health care access for underserved and vulnerable populations across 6 health domains, with mental health accounting for 6 of 18 studies. Mental health applications in Hazhir et al [20] included counseling and education, stress-management support, psychological distress screening, psychological first aid, grief support, and referral. Findings from both reviews were therefore treated as relevant to access and service delivery and were not assumed to transfer directly to all mental health services.

Adjacent and Indirect Evidence From Broader Health Care

Huynh et al [11] was classified as adjacent or partially direct because 23 of its 44 included reviews were mental health–focused. Wah [19] was also classified as adjacent or partially direct because mental health support was an explicit application domain, including counseling, stress management, and therapeutic support, although the review covered health care broadly and did not report a defensible count of mental health–specific included sources. In contrast, Chew and Achananuparp [24] and Feizollah et al [25] were classified as indirect yet relevant because mental health represented only a small subthread of broader health care evidence.

Across the 8 structured reviews, no direct review-level evidence was identified for workplace EAPs, student counseling services, workplace well-being services, or veteran well-being services; emergency-department mental health services were also essentially absent. Hazhir et al [20] added relevant evidence on access for underserved and vulnerable populations, although its findings span multiple health domains and did not constitute direct evidence for those specific mental health service settings. Findings from adjacent or indirect reviews were therefore treated as transferable hypotheses rather than demonstrated effects in those contexts.

RQ3: Applications Reported

Overview

Reported applications clustered across 4 broad areas, with substantial variation in evidence depth. The synthesis below is presented at the review level. Granular underlying-study examples and statistics reported by the included reviews were not treated as independently verified primary study evidence; their source review provenance is documented in Multimedia Appendix 5.

Clinical Decision Support, Screening, and Triage

Gu et al [23] identified 3 CDSS objectives across 21 studies: suicide risk reduction (n=1), diagnosis, treatment, or monitoring of mental health disorders (n=12), and real-time support for clinicians or counselors (n=8). Saif-Ur-Rahman et al [13] reported ML-based depression screening and tablet-based decision support for mental health and suicide risk in LMIC primary care. Huynh et al [11] reported diagnostic and triage accuracy as an outcome domain in 29 of its 44 included reviews. Hazhir et al [20] added mental health access applications, including psychological distress screening and referral, among underserved populations. Exact underlying-study performance figures reported by source reviews were treated as secondary-reported evidence and were not independently reanalyzed.

Conversational Service Delivery

Huynh et al [11] was the largest synthesis. Of the 23 reviews that reported mental health outcomes, 13 found AI-based conversational agents effective in improving or managing mental health, and 7 reported nonsignificant or mixed effects. The remaining 3 reviews were cited as reporting mental health outcomes but were not assigned to either group in the narrative synthesis of Huynh et al [11]; its extraction table described reductions in depression, anxiety, or stress for each of them. We reported the classification of Huynh et al [11] and did not reclassify these 3 reviews. Huynh et al [11] described the mental health evidence overall as mixed, with benefits reported more consistently in the short term than at later follow-up, and identified addiction-related applications as the only domain with consistently positive outcomes; safety signals included symptom worsening in one underlying chatbot study. Dwyer et al [22] reported text-based counseling outcomes alongside anonymity and convenience as engagement facilitators. Wah [19] described hybrid chatbot applications spanning mental health support, counseling, stress management, triage, follow-up, and patient engagement; its heterogeneous headline effectiveness claims were interpreted cautiously because the quantitative synthesis method was insufficiently specified. Hazhir et al [20] identified counseling, education, stress-management support, psychological first aid, grief support, and referral among the 6 mental health–focused studies in its 18 studies on health care access.

Engagement, Acceptability, and Adherence

Huynh et al [11] reported short-term engagement gains that were not consistently sustained at 3 to 6 months, while Saif-Ur-Rahman et al [13] reported improvements in clinic attendance, follow-up, and medication adherence in several mHealth studies. Wah [19] emphasized patient engagement, trust, and the proposed value of human-in-the-loop chatbot design. In Hazhir et al [20], only 9 of the 18 studies formally evaluated their chatbots; acceptability (n=6) and feasibility (n=5) were the most common evaluation outcomes, followed by satisfaction, usability, and technological performance.

Service-Level Analytics, Access, and Operational Improvement

Feizollah et al [25] was the principal source for NLP-based interpretation of unstructured patient feedback and reported using NLP to identify recurrent service issues and prioritize improvement themes, although it also identified a research-to-practice gap. Saif-Ur-Rahman et al [13] reported operational improvements in record-keeping, waiting times, appointment scheduling, facility crowding, and client satisfaction. Wah [19] reported hybrid chatbot applications in appointment and follow-up support, remote monitoring, workflow efficiency, and waiting-time reduction, although these heterogeneous claims were treated as secondary-reported findings rather than pooled evidence. Hazhir et al [20] shifted the service-level emphasis toward access: simple text-based interactions, familiar low-data platforms, and local-language adaptation were described as approaches to reducing linguistic, literacy, connectivity, and resource barriers.

Risk detection, client segmentation, and service planning were reported unevenly. Risk detection appeared most concretely in CDSS-based suicide risk identification [23], language-based suicide risk work [22], and mental health screening/referral applications within broader primary care or underserved-population reviews [13,20]. Client segmentation was largely absent from the synthesized evidence. Service-planning applications were limited mainly to operational workflow, access, appointment, and demand-management functions rather than dedicated BI or population-segmentation systems.

RQ4: Reported Benefits, Insights, Challenges, and Barriers to Implementation

Reported benefits, insights, challenges, and barriers were synthesized narratively below under 4 headings: benefits and enablers; challenges and barriers; the integration gap; and ethical, safety, and equity considerations.

Benefits and Enablers

On-demand availability, accessibility, ease of use, patient engagement, and selected efficiency gains were reported across several included reviews [11,13,19,20,23,24]. Service-delivery benefits, including scheduling, record-keeping, waiting-time reduction, access, and client satisfaction, were reported more consistently than sustained clinical benefits,

although the underlying evidence was heterogeneous and often broader than mental health service delivery. Dwyer et al [22] emphasized anonymity and convenience in text-based mental health communication; Wah [19] emphasized hybrid human-AI oversight, workflow integration, and personalization; and Hazhir et al [20] highlighted simple text interfaces, familiar low-data platforms, local-language adaptation, and immediate low-cost access for underserved populations. These findings were interpreted as reported enablers rather than established causal effects.

Challenges and Barriers

Privacy, trust, safety, workflow integration, and infrastructure constraints recurred across the evidence base. Chew and Achananuparp [24] reported privacy, patient-safety, and automation concerns across its included studies. Gu et al [23] reported alert fatigue, slow CDSS performance, data-entry burden, and EHR-integration problems, while Saif-Ur-Rahman et al [13] described resource, connectivity, language, digital-literacy, and workflow constraints in LMIC primary care. Wah [19] added concerns about trust in AI advice, data security, accountability, algorithmic bias, interoperability, cost, cultural adaptability, and the need for human expertise in complex cases. Hazhir et al [20] similarly identified weak infrastructure, limited internet access, low digital and health literacy, language barriers, privacy/security concerns, development costs, and inconsistent evaluation as barriers to equitable chatbot deployment.

The Integration Gap

Two reviews independently identify a research-to-practice gap as a central finding. Feizollah et al [25] noted that academic rather than translational impact had been the driving force behind research in NLP-based patient feedback work and that few studies describe concrete clinical impact from NLP-derived insights. Gu et al [23] identified the absence of a holistic plan for the full CDSS development cycle as a key finding. Most included studies covered only a subset of the usability-software development life cycle (U-SDLC) phases. The maintenance phase was particularly poorly covered, with 5 development activities at zero across all 21 included studies.

Ethical, safety, and equity considerations were synthesized unevenly across the 8 reviews. Chew and Achananuparp [24] and Huynh et al [11] identified privacy, safety, bias, hallucination, and symptom-worsening concerns, while Wah [19] emphasized trust, accountability, privacy/security, bias, and cultural adaptation. Hazhir et al [20] contributed the clearest additional access-equity perspective by focusing explicitly on underserved and vulnerable populations, including linguistic, geographic, literacy, and resource barriers; however, only half of its included studies formally evaluated their chatbot, and the long-term impact remained unclear. Saif-Ur-Rahman et al [13] added LMIC-specific digital divide and infrastructure concerns.

Primary-Study Overlap and CCA Analysis

No verified overlap was found among the included studies that could be identified from the included study/source lists (Table 4). In Tier 1, the 7 primary-level reviews contributed N=264 included study/source occurrences and r=264 unique units across C=7 reviews, giving CCA=0.00%. Nine of the 52 studies included by Feizollah et al [25] were listed only in a supplementary file and could not be identified; if all 9 duplicated a study in another Tier 1 review, r would fall to 255 and the CCA would be 0.59%, which would remain in the slight category. Wah [19] reported 29 included sources of mixed type rather than a wholly primary study set; because none overlapped with another Tier 1 review, excluding its nonprimary sources would reduce N and r equally and would not change the zero-overlap numerator. In Tier 2, the 2 reviews classified as directly relevant to mental health service delivery, Dwyer et al [22] and Gu et al [23], contributed N=91 occurrences and r=91 unique studies, giving a CCA of 0.00%. In Tier 3, none of the 7 standalone reviews was among the 44 constituent reviews of Huynh et al [11] (0 of 7); this review-level result was reported as a direct count rather than as a CCA. Two cross-tier relationships were identified (Milne-Ives et al [26] and Aggarwal et al [27]); because these involved different units of analysis, they were reported separately rather than being counted in any tier's CCA. Primary studies nested inside all 44 constituent reviews of Huynh et al [11] were not fully disaggregated. Full matrices and calculations are provided in Multimedia Appendix 7.

Table 4. Primary study CCA[a] and review-level overlap across the 3 analytical tiers.

Analytical tier	Reviews included (C)[b]	Total occurrences (N)	Unique units (r)	Shared units	CCA (%)	Interpretation
Tier 1: full primary-level evidence base	7	264	264	0	0.00% (≤0.59% worst case)	No verified overlap among identifiable studies; slight category (<5%)
Tier 2: direct mental health service subset	2	91	91	0	0.00%	No verified overlap; slight category (<5%)
Tier 3: review-level overlap check	—[c]	—	—	0	Not calculated	No review-level overlap (0 of 7 standalone reviews among 44 constituent reviews); direct count

[a]CCA: corrected covered area.

[b]The columns report counts and one derived index: C is the number of reviews included in each tier; N is the total number of included study or source occurrences across those reviews; r is the number of unique studies or sources after deduplication; shared units is the number of studies or

sources appearing in more than 1 review. CCA (corrected covered area) is calculated as CCA=(N–r)/(r×(C–1)) [21] and expressed as a percentage, with values below 5% interpreted as slight overlap.
[c]Not applicable.

Discussion

Principal Findings

This umbrella review synthesized 8 structured reviews on data-driven methods relevant to mental health and well-being service delivery, published between 2021 and 2026. Five main observations emerged.

First, the methods landscape was heavily weighted toward AI and ML approaches (conversational agents, NLP, deep learning, and LLMs), while BI and operational data visualization as distinct technology categories were thinly represented in the review-level literature retrieved by our searches.

Second, and consistent with the service-delivery orientation of this review, service-level outcomes (efficiency, accessibility, satisfaction, and engagement) were reported more consistently and more positively than clinical outcomes (symptom change, behavior change, sustained effect) within the included reviews. This described the pattern of reporting across a small and heterogeneous evidence base rather than establishing stronger service-level effectiveness. Service-level outcomes were nonetheless reported less often than technical performance measures, such as algorithm accuracy (see Comparison With Prior Work).

Third, in the conversational-agent evidence synthesized by Huynh et al [11], short-term benefits were reported more often than long-term benefits and were not reliably sustained at 3- to 6-month follow-up; this review describes the mental health evidence overall as mixed.

Fourth, evidence directness varies materially across the 8-review set. Dwyer et al [22] and Gu et al [23] provided the 2 direct sources for mental health service delivery. Huynh et al [11], Saif-Ur-Rahman et al [13], Wah [19], and Hazhir et al [20] provided adjacent or partially direct evidence because mental health is a substantial or explicit component of broader health care, primary care, or access-focused syntheses. Chew and Achananuparp [24] and Feizollah et al [25] were indirect yet relevant. Consequently, cross-cutting implementation and service findings from adjacent and indirect sources were treated as potentially informative rather than as demonstrated effects in specific mental health service settings.

Fifth, a research-to-practice gap is a recurring cross-cutting finding, identified independently by Gu et al [23] and Feizollah et al [25].

Comparison With Prior Work

Huynh et al [11] is itself an umbrella review; its scope, however, is limited to conversational agents and clinical health-related outcomes across health care. This review used Huynh et al [11] as one of 8 structured inputs and places its findings alongside clinical decision support, NLP-based patient-feedback analysis, text-based mental health communication, AI adoption, LMIC digital health, hybrid health care chatbots, and chatbot-enabled health care access for underserved populations. This broader service-delivery framing permitted comparison across technology and service-context families while retaining explicit evidence-directness distinctions.

A scoping review by Woo and Jung [28], published online on April 3, 2026, after the retained January 27, 2026, evidence cutoff, was retrieved by the supplementary search. It is cited here only as contextual literature and was not included in the formal synthesis, critical appraisal, evidence tables, or CCA calculations. Its publication reflected the pace of review activity in this area and did not alter the prespecified evidence window of the formal synthesis.

The synthesis also aligned with earlier method-focused reviews by Shatte et al [6] and Le Glaz et al [7], which concluded that methodological promise had outpaced clinical translation. The present review extended that observation to service delivery: operational outcomes, such as waiting time, caseload, triage, and demand, were reported less often than technical performance.

Methods Landscape: Where the Review-Level Evidence Concentrates

Dedicated BI, dashboard, and operational-visualization syntheses were not identified. Because the search required all 3 concept blocks, this finding applies within the searches conducted rather than proof of absence from the wider literature.

Within the AI/ML literature, methods ranged from rule-based conversational systems and supervised learning to topic modeling, sentiment analysis, and LLMs. This heterogeneity limits meaningful quantitative resynthesis at the umbrella review level [11].

The label “AI or digital health” did not correspond to substantive ML in most studies within the included LMIC primary-care review. Saif-Ur-Rahman et al [13] reported that only 1 of 48 studies used ML, while the remainder primarily involved mHealth or other digital health technologies. This observation applies only to that review.

Service Contexts and the Helpline-Specific Evidence Gap

Mental health helpline services were the explicit focus of only one review [23], and even this direct source had a broader effective scope than its title suggested. For helpline operators and service planners, the evidence should therefore be read as limited rather than as a mature helpline-specific synthesis. Some implementation findings were derived from adjacent primary care, telepsychiatry, online therapy,

health care chatbots, and access-focused contexts [13,19,20]. Co-design, EHR integration, human oversight, alert-fatigue management, linguistic adaptation, and stepwise rollout may be relevant to helplines; however, transferability remains an inference rather than a directly demonstrated effect.

Beyond helplines, workplace EAPs, student counseling, workplace well-being, and veteran well-being were not represented by direct included review-level evidence. Hazhir et al [20] broadened the evidence on underserved and vulnerable populations and Wah [19] broadened the chatbot implementation literature; neither was a direct synthesis of these specific mental health service settings. Dedicated primary studies and context-specific syntheses are therefore needed before transfer can be assumed.

Applications and Outcomes: Service-Level Signal, Clinical-Level Caution

As noted above, service-level outcomes were reported more consistently than clinical outcomes. Given the small, heterogeneous, and partly indirect evidence base, this reflects consistency of reporting rather than established service-level effectiveness, with 3 implications.

First, the included reviews do not provide a sufficiently direct or homogeneous evidence base to conclude that BI, visualization, or AI are established as effective at the service-delivery level. Second, where clinical effects are reported, they are predominantly short-term, particularly in the conversational-agent evidence synthesized by Huynh et al [11]. Third, safety and equity signals included symptom worsening, LLM hallucination and bias, inaccurate information, privacy and trust concerns, language and literacy barriers, and unequal digital access [11,19,20,24]. These issues are reported across several reviews and are not yet synthesized with sufficient consistency or long-term follow-up to support strong service-wide conclusions.

RQ5: Research Gaps and Recommended Directions for Future Inquiry

Overview

Five priorities for future research emerged from this synthesis.

Helpline-Specific Implementation Evidence

Primary evaluation studies of BI, visualization, and AI applications in mental health helpline operations are needed before higher-order synthesis is possible. These include evaluations of CDSS deployed in real-time counselor decision-making, visual analytics dashboards for service planners, and predictive models for demand and risk in helpline contexts. Service-analytics platforms of this kind should be designed with evaluation embedded from the outset, in line with implementation science methods [23].

Service-Delivery-Focused Evaluation

Existing reviews concentrated on technology performance (algorithm accuracy, sentiment-classification F_1, and chatbot engagement metrics) rather than on service-delivery impact (waiting times, caseload management, dropout reduction, and equity of access). Future primary studies and reviews should foreground service-delivery outcomes and report them at parity with clinical and technical outcomes.

Longer Follow-up and Treatment-Effect Durability

The finding reported by Huynh et al [11], that short-term benefits of AI-based conversational agents for mental health were not consistently sustained at 3- to 6-month follow-up, indicates that effect durability is a first-order RQ. Future trials and reviews should report follow-up beyond 3 months by default and address the mechanisms of effect decay.

Explicit Treatment of Safety, Ethics, and Equity

Beyond the safety signals noted above [11,19,24], Hazhir et al [20] showed that language, literacy, connectivity, cultural fit, and resource constraints were central to equitable access for underserved populations. Future reviews should treat safety and equity as primary synthesis objectives, use explicit adverse-outcome frameworks, and examine differential access and performance across demographic, linguistic, and service-context groups.

Frameworks Linking Design, Implementation, and Evaluation

Within the formal evidence base, Gu et al [23] identified the absence of a holistic development and implementation plan across the full CDSS lifecycle. Outside the formal umbrella synthesis, Lee et al [12] provided a complementary conceptual digital-therapeutics analytics framework, including just-in-time adaptive intervention (JITAI) and engagement-adherence distinctions. This contextual source is used to enrich interpretation rather than as formal review-level evidence. Bridging the design-implementation-evaluation gap will require collaboration between health services research, implementation science, software engineering, and clinical mental health practice.

Implications for Practice, Policy, and Implementation

For service operators and planners, the included reviews suggested potentially useful service-delivery signals for digital and AI-based interventions, including accessibility, workflow, engagement, and selected operational outcomes. However, applicability depends strongly on evidence directness: mental health service-specific conclusions are anchored principally in the 2 direct reviews, while findings from the 4 adjacent and 2 indirect health care syntheses should be treated as transferable hypotheses rather than established effects in a particular service context. Clinical-effectiveness evidence remains limited and predominantly short-term.

Technical feasibility or promising retrospective performance alone should not be treated as sufficient justification for deployment in mental health services. AI, BI, dashboard, and decision-support tools should be introduced

with prospective service-level evaluation, predefined safety monitoring and escalation pathways, equity assessment across relevant demographic, language, and access groups, and explicit workflow-integration planning with service users and staff. Staged implementation, human oversight, monitoring for unintended consequences and alert burden, and predefined criteria for modification or withdrawal should accompany routine deployment, particularly when the evidence is indirect or context transfer is assumed.

Strengths and Limitations of This Umbrella Review

Strengths

A time-stamped OSF draft of the scoping review protocol existed from December 2, 2025, before the original database searches on January 27, 2026, and documented the original search strategy, information sources, and publication window. The interdisciplinary original search combined multidisciplinary, psychological, and computing databases, and revision-stage PubMed and Embase searches were added transparently to address reviewer concerns while retaining the original evidence cutoff. Title/abstract and full-text screening were conducted independently by 2 reviewers in both search phases: SA and EE for the original search and SA and GC for the supplementary PubMed/Embase records. Review-team discussion and final eligibility adjudication were available where required. JBI critical appraisal was conducted independently by 2 reviewers with third-reviewer adjudication, and lead data extraction was independently cross-checked against source reviews. The formal synthesis was restricted to structured reviews, evidence directness was explicit, and overlap was reassessed with citation-level verification, using a formal CCA at 2 primary-study tiers and a review-level overlap check.

Limitations

The review had eight main limitations. (1) Search sensitivity for BI and visualization could not be assumed to be equivalent to sensitivity for AI/ML. The development-stage reference check used 5 known primary studies from the original scoping review design rather than a review-level BI/dashboard validation set; all 3 Boolean concept blocks were mandatory, and the ACM Digital Library search was restricted to the abstract field. The absence of dedicated BI or visualization syntheses should therefore be interpreted only within the retrieved review-level evidence. (2) PubMed and Embase were added during revision (September 22, 2026) rather than in the original search. The January 27, 2026, evidence cutoff was enforced during screening, and the supplementary searches did not constitute a synchronized rerun of all databases. Wah [19] was excluded at original title-and-abstract screening and included after supplementary screening, which shows that title-and-abstract judgments for broad health care reviews with a secondary mental health component could differ between reviewer pairs; other reviews of this kind may have been missed. CINAHL and the Cochrane Library were not searched directly. (3) Eligibility was restricted to English-language peer-reviewed publications from 2021 to the retained January 27, 2026, cutoff, excluding gray literature, preprints, non-English reviews, and earlier syntheses. (4) The formal evidence base remained small and heterogeneous (8 structured reviews). Only 2 reviews were direct to mental health service delivery, 4 were adjacent or partially direct, and 2 were indirect yet relevant; only Gu et al [23] was explicitly anchored in mental health helpline services, and no direct included review-level evidence was identified for EAPs, student counseling, workplace well-being, or veteran services. (5) The umbrella review by Huynh et al [11] contained a per-item appraisal table that is incompletely legible in the published article, limiting independent verification of its own appraisal. (6) Extraction necessarily relied on what source reviews reported. Primary-study-level examples in the supplementary materials were not independently verified against the primary publications and may inherit source-review reporting limitations; Wah [19] also reported 29 included sources of mixed type and does not clearly specify the quantitative synthesis underlying several headline estimates. (7) The time-stamped December 2025 OSF draft was a scoping review protocol. The umbrella review design was adopted and formally registered on May 6, 2026, after the original January searches; the design was therefore not prospectively registered before searching. Two nonsystematic publications included in the originally submitted version were removed from the formal evidence base when eligibility was standardized to structured syntheses only. (8) The CCA was 0.00% at both primary-study tiers among the studies that could be identified. Nine of the 52 studies included by Feizollah et al [25] could not be identified from the published article; if all 9 were duplicates, the Tier 1 CCA would be 0.59%. Zero verified overlap should not be interpreted as proof that the underlying evidence bases are independent in every respect. Primary studies nested within all 44 constituent reviews of Huynh et al [11] were not fully disaggregated, and cross-tier relationships were reported rather than combined because they involve different units of analysis (Multimedia Appendix 7).

Conclusions

Review-level evidence on data-driven methods relevant to mental health and well-being service delivery was uneven and varies substantially in directness. In terms of the RQs, the data-driven methods, tools, algorithms, and datasets reported across the included reviews (RQ1) were dominated by AI and ML, including conversational agents, NLP, LLMs, and CDSS; dedicated BI and operational data visualization syntheses were sparse within the retrieved evidence, and dataset-level reporting was limited. The mental health service contexts in which these methods have been applied (RQ2) were text-based counseling and helplines, LMIC primary care, health care access for underserved populations, and broader health care settings; 2 reviews were direct to mental health service delivery, 4 were adjacent or partially direct, and 2 were indirect yet relevant. Only one review focused specifically on helplines, and no direct review-level evidence was identified for EAPs, student counseling, workplace well-being, or veteran services. The applications (RQ3) were clinical decision support, screening,

and triage; conversational service delivery; engagement and adherence support; and service-level analytics and access improvement, whereas client segmentation and BI-based service planning were largely absent. Among the benefits, insights, challenges, and barriers to implementation (RQ4), service-level benefits, such as accessibility, efficiency, and satisfaction, were reported more consistently than sustained clinical benefits, and recurring barriers included privacy and trust concerns, safety risks such as symptom worsening and LLM hallucination, workflow and EHR integration problems, infrastructure and digital-literacy constraints, and a research-to-practice gap. The research gaps and directions for future inquiry (RQ5) were identified as helpline-specific implementation evidence, service-delivery-focused evaluation, longer follow-up of effect durability, explicit treatment of safety, ethics, and equity, and frameworks linking design, implementation, and evaluation. Because the evidence base is small, heterogeneous, and partly indirect, and the mandatory 3-concept search may have under-retrieved BI- and visualization-focused reviews, these findings should guide prospective, context-specific evaluation rather than be generalized across service settings.

Acknowledgments

The authors thank Mary Rose, Librarian, for assistance in refining search terms.
All substantive methodological decisions, data extraction, critical appraisal, interpretation, and final manuscript content were undertaken or verified by the authors, who take full responsibility for the work.

Generative AI Use

Generative AI tools (ChatGPT and Copilot) were used for language refinement, formatting, and consistency checking.

Funding

The authors declared no specific financial support was received for this work. SA is employed as an Associate on a Knowledge Transfer Partnership (KTP) between Ulster University and Inspire Wellbeing, funded by Innovate UK, and the review was conducted as part of that role. Neither Inspire Wellbeing nor Innovate UK participated in framing the review question, determining eligibility criteria, screening or appraisal decisions, reviewing interim findings, drafting or revising the manuscript, or decisions concerning submission and publication.

Data Availability

The structured data-extraction workbook, item-level JBI appraisal records, PRISMA materials, provenance table, full evidence matrix, and citation-overlap/CCA materials are provided with this article as Multimedia Appendices 1-7 and Checklist 1. The registered protocol is available through the OSF registration [17].

Authors' Contributions

Conceptualization: SA, RB, ZT
Data curation: SA
Formal analysis: SA
Investigation: SA, GC, EE, DB
Methodology: SA, RB, MM, EE
Project administration: SA
Supervision: RB, ZT, GC, EE
Validation: GC, MM, EE
Writing – original draft: SA
Writing – review & editing: SA, RB, ZT, GC, MM, EE, DB, KM

Conflicts of Interest

SA is employed through the Knowledge Transfer Partnership (KTP) described under Funding, and this review was conducted in part to inform the design of a service-analytics platform for Inspire Wellbeing. A time-stamped OSF registration draft was created on December 2, 2025, as a scoping-review protocol before the final database searches; the protocol was subsequently revised to an umbrella review design and formally registered on May 6, 2026, after searching had been completed. The registration is therefore not presented as a prospective safeguard for the umbrella review search stage. To reduce the possibility that the applied KTP context influenced eligibility or interpretation, screening was conducted independently by 2 reviewers, critical appraisal was conducted independently by 2 reviewers with third-reviewer adjudication, data extraction and evidence-directness classifications were independently verified by a second reviewer, and final eligibility conflicts were adjudicated by MM. No other competing interests are declared.

Multimedia Appendix 1

Full electronic search strategies.
[DOCX File (Microsoft Word File), 20 KB-Multimedia Appendix 1]

Multimedia Appendix 2

PsycINFO controlled-vocabulary and free-text concept blocks.
[DOCX File (Microsoft Word File), 160 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Full data extraction spreadsheet.
[XLSX File (Microsoft Excel File), 19 KB-Multimedia Appendix 3]

Multimedia Appendix 4

JBI critical appraisal item-by-item records.
[XLSX File (Microsoft Excel File), 20 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Provenance of underlying-study examples and statistics reported by included reviews.
[DOCX File (Microsoft Word File), 43 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Full review-level evidence matrix of methods, service contexts, applications, and evidence directness.
[DOCX File (Microsoft Word File), 42 KB-Multimedia Appendix 6]

Multimedia Appendix 7

Primary-study citation matrix and multitiered corrected covered area analysis.
[DOCX File (Microsoft Word File), 59 KB-Multimedia Appendix 7]

Checklist 1

PRISMA 2020 checklist.
[DOCX File (Microsoft Word File), 23 KB-Checklist 1]

References

1. World Health Organization. World mental health report: transforming mental health for all. Geneva: WHO; 2022. URL: https://www.who.int/publications/i/item/9789240049338 [Accessed 2026-11-05]
2. Cost of living and depression in adults, Great Britain: 29 September to 23 October 2022. Office for National Statistics. 2022. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/mentalhealth/articles/costoflivinganddepressioninadultsgreatbritain/29septemberto23october2022 [Accessed 2026-05-11]
3. Hopkin G, Coole H, Edelmann F, et al. Toward a new conceptual framework for digital mental health technologies: scoping review. JMIR Ment Health. Feb 19, 2025;12:e63484. [doi: 10.2196/63484] [Medline: 39969824]
4. Wills MJ. Decisions through data: analytics in healthcare. J Healthc Manag. 2014;59(4):254-262. [doi: 10.1097/00115514-201407000-00005] [Medline: 25154123]
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. Jan 2019;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
6. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med. Jul 2019;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
7. Le Glaz A, Haralambous Y, Kim-Dufor DH, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res. May 4, 2021;23(5):e15708. [doi: 10.2196/15708] [Medline: 33944788]
8. Ehtemam H, Sadeghi Esfahlani S, Sanaei A, et al. Role of machine learning algorithms in suicide risk prediction: a systematic review-meta analysis of clinical studies. BMC Med Inform Decis Mak. May 27, 2024;24(1):138. [doi: 10.1186/s12911-024-02524-0] [Medline: 38802823]
9. Balcombe L, De Leo D. Evaluation of the use of digital mental health platforms and interventions: scoping review. Int J Environ Res Public Health. 2023;20(1):362. [doi: 10.3390/ijerph20010362]
10. Zhong S, Yang X, Pan Z, et al. The usability, feasibility, acceptability, and efficacy of digital mental health services in the COVID-19 pandemic: scoping review, systematic review, and meta-analysis. JMIR Public Health Surveill. Feb 13, 2023;9:e43730. [doi: 10.2196/43730] [Medline: 36634261]
11. Huynh AL, Roy TJ, Jackson KN, Lee AG, Liaw W, Hossain MM. Applications of artificial intelligence-based conversational agents in healthcare: a systematic umbrella review. Int J Med Inform. Mar 1, 2026;207:106204. [doi: 10.1016/j.ijmedinf.2025.106204] [Medline: 41337874]
12. Lee U, Jung G, Ma EY, et al. Toward data-driven digital therapeutics analytics: literature review and research directions. IEEE/CAA J Autom Sinica. 2023;10(1):42-66. [doi: 10.1109/JAS.2023.123015]

13. Saif-Ur-Rahman KM, Islam MS, Alaboson J, et al. Artificial intelligence and digital health in improving primary health care service delivery in LMICs: a systematic review. J Evid Based Med. Sep 2023;16(3):303-320. [doi: 10.1111/jebm.12547] [Medline: 37691394]
14. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc. Sep 2015;13(3):132-140. [doi: 10.1097/XEB.0000000000000055] [Medline: 26360830]
15. Pollock M, Fernandes RM, Becker LA, Pieper D, Hartling L. Chapter V: overviews of reviews. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions, Version 6.5. Cochrane; 2024. URL: https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-v [Accessed 2026-11-11]
16. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. Mar 29, 2021;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
17. Ahmad S, Bond R, Tariq Z, Cameron G, Mulvenna MD, Ennis E, et al. Umbrella review: BI, data analytics, and ML in mental health and well-being services. OSF Registries. Preprint posted online on May 6, 2026. [doi: 10.17605/OSF.IO/8NQBC]
18. Mumtaz H, Riaz MH, Wajid H, et al. Current challenges and potential solutions to the use of digital health technologies in evidence generation: a narrative review. Front Digit Health. 2023;5:1203945. [doi: 10.3389/fdgth.2023.1203945] [Medline: 37840685]
19. Wah JNK. Revolutionizing e-health: the transformative role of AI-powered hybrid chatbots in healthcare solutions. Front Public Health. 2025;13:1530799. [doi: 10.3389/fpubh.2025.1530799] [Medline: 40017541]
20. Hazhir S, Langarizadeh M, Bahariniya S, Valizadeh Laktarashi H, Hami A, Fatemi Aghda SA. The development and use of chatbots in enhancing health care access for underserved and vulnerable populations: a scoping review. BMC Public Health. 2026;26(1):410. [doi: 10.1186/s12889-025-25814-2]
21. Pieper D, Antoine SL, Mathes T, Neugebauer EAM, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. J Clin Epidemiol. Apr 2014;67(4):368-375. [doi: 10.1016/j.jclinepi.2013.11.007] [Medline: 24581293]
22. Dwyer A, de Almeida Neto A, Estival D, Li W, Lam-Cassettari C, Antoniou M. Suitability of text-based communications for the delivery of psychological therapeutic services to rural and remote communities: scoping review. JMIR Ment Health. Feb 24, 2021;8(2):e19478. [doi: 10.2196/19478] [Medline: 33625373]
23. Gu Y, Andargoli AE, Mackelprang JL, Meyer D. Design and implementation of clinical decision support systems in mental health helpline services: a systematic review. Int J Med Inform. Jun 2024;186:105416. [doi: 10.1016/j.ijmedinf.2024.105416] [Medline: 38552266]
24. Chew HSJ, Achananuparp P. Perceptions and needs of artificial intelligence in health care to increase adoption: scoping review. J Med Internet Res. Jan 14, 2022;24(1):e32939. [doi: 10.2196/32939] [Medline: 35029538]
25. Feizollah A, Lin CY, O'Malley L, Thompson W, Listl S, Byrne M. The use of natural language processing to interpret unstructured patient feedback on health services: scoping review. J Med Internet Res. Aug 14, 2025;27:e72853. [doi: 10.2196/72853] [Medline: 40811803]
26. Milne-Ives M, de Cock C, Lim E, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res. Oct 22, 2020;22(10):e20346. [doi: 10.2196/20346] [Medline: 33090118]
27. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res. Feb 24, 2023;25:e40789. [doi: 10.2196/40789] [Medline: 36826990]
28. Woo Y, Jung K. Artificial intelligence–driven tools in mental health service delivery: a scoping review. Healthcare (Basel). Apr 3, 2026;14(7):943. [doi: 10.3390/healthcare14070943] [Medline: 41975946]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
BI: business intelligence
CCA: corrected covered area
CDSS: clinical decision support system
EAP: employee assistance program
EHR: electronic health record
IoT: Internet of Things
IVR: interactive voice response
JBI: Joanna Briggs Institute
JITAI: just-in-time adaptive intervention
LDA: latent Dirichlet allocation

LIWC: linguistic inquiry and word count
LLM: large language model
LMIC: low- and middle-income country
mHealth: mobile health
ML: machine learning
NA: not applicable
NLP: natural language processing
OSF: Open Science Framework
PCC: population-concept-context
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question
SVM: support vector machine
U-SDLC: usability-software development life cycle
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by Matthew Balcarras; peer-reviewed by Janet McMordie, Peng Liu; submitted 02.Jul.2026; initial comments to author 26.Aug.2026; final revised version received 16.Oct.2026; accepted 16.Oct.2026; published 25.Nov.2026

Please cite as:
Ahmad S, Bond R, Tariq Z, Cameron G, Mulvenna M, Ennis E, Miller K, Buckley D
Data-Driven Methods Supporting Mental Health and Well-Being Service Delivery: Umbrella Review
J Med Internet Res 2026;28:e106104
URL: https://www.jmir.org/2026/1/e106104
doi: 10.2196/106104

© Salman Ahmad, Raymond Bond, Zeeshan Tariq, Gillian Cameron, Maurice Mulvenna, Edel Ennis, Kristel Miller, Daniel Buckley. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.